

**NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH**  
**THE TETRAPYRROLE PATHWAY**

**CROSS REFERENCE TO RELATED APPLICATIONS**

This application claims priority under 35 U.S.C §119(e) and/or 35 U.S.C §120 of applications No. 60/067000 filed November 24, 1997, No. 60/069472 filed December 9, 1997, No. 60/072,027 filed January 21, 1998, No. 60/074,201 filed February 10, 1998, No. 60/074282 filed February 10, 1998, No. 60/074280 filed February 10, 1998, No. 60/074281 filed February 10, 1998, No. 60/074566 filed February 12, 1998, No. 60/074567 filed February 12, 1998, No. 60/074565 filed February 12, 1998, No. 60/075462 filed February 19, 1998, No. 60/075459 filed February 19, 1998, No. 60/075461 filed February 19, 1998, No. 60/075464 filed February 19, 1998, No. 60/075460 filed February 19, 1998, No. 60/075463 filed February 19, 1998, No. 60/077231 filed March 9, 1998, No. 60/077229 filed March 9, 1998, No. 60/077230 filed March 9, 1998, No. 60/078368 filed March 18, 1998, No. 60/080844 filed April 7, 1998, No. 60/083067 filed April 27, 1998, No. 60/083387 filed April 29, 1998, No. 60/083388 filed April 29, 1998, No. 60/083389 filed April 29, 1998, No. 60/085224 filed May 13, 1998, No. 60/085223 filed May 13, 1998, No. 60/085222 filed May 13, 1998, No. 60/086186 filed May 21, 1998, No. 60/086187 filed May 21, 1998, No. 60/086185 filed May 21, 1998, No. 60/086184 filed May 21, 1998, No. 60/086183 filed May 21, 1998, No. 60/086188 filed May 21, 1998, No. 60/089,524 filed June 16, 1998, No. 60/089,810 filed June 18, 1998, No. 60/089,814 filed June 18, 1998, No. 60/091,035 filed June 29, 1998, No. 60/091,405 filed June 30, 1998, "Nucleic Acid Molecules and Other Molecules Associated with the Plant Sugar and Nitrogen Transporters Pathway" docket No. 38-21(15412)A filed June 30, 1998, No. 60/099670 filed September 9, 1998, No. 60/099697 filed September 9, 1998, No. 60/100674 filed September 16, 1998, No. 60/100672

filed September 16, 1998, No. 60/101130 filed September 21, 1998, No. 60/101,508 filed September 22, 1998, No. 60/101344 filed September 22, 1998, No. 60/101347 filed September 22, 1998, No. 60/101343 filed September 22, 1998, No. 60/104,128 filed October 13, 1998, No. 60/104,127 filed October 13, 1998, No. 60/109,018 filed November 18, 1998, No. 60/108,996 filed November 18, 1998, "Nucleic Acid Molecules and Other Molecules Associated With Plants" docket No. 38-21(15075)B filed November 24, 1998, No. 09/210,297 filed December 8, 1998, "Nucleic Acid Molecules and Other Molecules Associated with Plants" docket No. 38-21(15668)A filed December 11, 1998 and No. 60/113,224 filed December 22, 1998, all of which are herein incorporated by reference in their entirety.

### **FIELD OF THE INVENTION**

The present invention is in the field of plant biochemistry. More specifically the invention relates to nucleic acid sequences from plant cells, in particular, nucleic acid sequences from maize and soybean plants associated with the tetrapyrrole pathway in plants. The invention encompasses nucleic acid molecules that encode proteins and fragments of proteins. In addition, the invention also encompasses proteins and fragments of proteins so encoded and antibodies capable of binding these proteins or fragments. The invention also relates to methods of using the nucleic acid molecules, proteins and fragments of proteins and antibodies, for example for genome mapping, gene identification and analysis, plant breeding, preparation of constructs for use in plant gene expression and transgenic plants.

### **BACKGROUND OF THE INVENTION**

#### **I. BIOSYNTHESIS OF TETRAPYRROLES**

The biosynthesis of tetrapyrroles such as heme and chlorophyll as well as a number of other tetrapyrroles such as siroheme, the cofactor for sulfite and nitrite reductases, cobalamin

(vitamin B12), and the chromophore of phytochrome, can be subdivided into three major phases; ALA synthesis, porphyrin ring synthesis and synthesis of final products. The pathway is conserved among species except for the synthesis of 5-aminolevulinate, also known as 5-aminolevulinic acid ("ALA") (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995), both of which are herein incorporated by reference).

The first phase of the biosynthesis of tetrapyrroles, such as heme and chlorophyll, is the synthesis of ALA. Yeast, fungi, mammals and some bacteria (the  $\alpha$ -group of proteobacteria or purple bacteria, *e.g.* *Bradyrhizobium japonicum* and *Rhodobacter capsulatus*) biosynthesize tetrapyrroles via the single step four-carbon (C4), or Shemin pathway. In this pathway ALA synthase (E.C. 2.3.1.37) catalyzes the condensation of glycine with succinyl-CoA to generate ALA.

Plants, green algae, cyanobacteria, most eubacteria (*e.g.* *E. coli* and *Bacillus subtilis*), and archaeobacteria biosynthesize ALA via the three-step five-carbon ("C5") pathway, which includes glutamyl-tRNA synthetase ("GluRS"), glutamyl-tRNA reductase ("GluTR") and glutamate-1-semialdehyde aminotransferase ("GSA-AT"). In plants and algae, the C5 pathway is localized in the chloroplast. The formation of ALA via the C5 pathway is reported to be the rate-limiting step in the biosynthesis of heme and chlorophyll (Kumar *et al.*, *Trends in Plant Science* 1:371-376 (1996); Tanaka *et al.*, *Plant Physiol.* 110:1223-30 (1996); Masuda *et al.*, *Plant Physiol. Biochem.* 34:11-16 (1996); Hungerer *et al.*, *J. Bacteriol.* 177:1435-43 (1995); Ilag *et al.*, *Plant Cell* 6:265-75 (1994), all of which are herein incorporated by reference in their entirety).

Chloroplastic GluRS (E.C. 6.1.1.17), also known as glutamate-tRNA ligase, converts glutamate to glutamyl-tRNA ("Glu-tRNA") activating the C-1 of glutamate in an ATP dependent reaction (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). Glu-tRNA is reported to be the first intermediate in the C5 pathway and it also reported to serve as a source of glutamate in protein biosynthesis. GluRS is a soluble plastid enzyme which has been isolated from higher plants (barley, wheat) and other organisms. Reported GluRS enzymes are homodimers encoded by a nuclear gene and synthesized in the cytoplasm and have a molecular weight of 54 kD (barley) and 56 kD (wheat).

GluTR, the first committed enzyme reported in heme and chlorophyll biosynthesis, catalyzes the NADPH dependent reduction of Glu-tRNA to glutamate 1-semialdehyde ("GSA") with the release of intact tRNA (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). GluTR is reported as the rate limiting step in ALA formation and is present only at low levels in all organisms examined (Masuda *et al.*, *Plant Physiol. Biochem.* 34:11-16 (1996); Schroeder *et al.*, *Biochem. J.* 281:843-50 (1992), the entirety of which is herein incorporated by reference; Masuda *et al.*, *Plant Cell Physiol.* 36:1237-43 (1995), the entirety of which is herein incorporated by reference). Plant GluTR is a soluble enzyme localized in plastids and encoded in the nucleus. GluTR has been reported to exist as a multimer of a single subunit. The purified barley enzyme has a molecular weight of 270 kD with a monomeric subunit size of 54 kD (Pontoppidan and Kannangara, *Eur. J. Biochem.* 225:529-37 (1994), the entirety of which is herein incorporated by reference). *Arabidopsis* and cucumber enzymes have similar subunit molecular weights (Tanaka *et al.*, *Plant Physiol.* 110:1223-30 (1996); Ilag *et al.*, *Plant Cell* 6:265-75 (1994); Kumar *et al.*, *Plant Mol. Biol.* 30:419-26 (1996), the entirety of which is herein incorporated by reference).



GluTR genes (also known as HEMA genes) have been cloned and the amino acid sequences determined for a number of sources including three higher plants; *Arabidopsis*, barley, and cucumber. The deduced amino acid sequence of GluTR from all sources exhibit about 60% overall similarity with stretches of amino acid identity. Barley, *Arabidopsis*, and cucumber show over 70% identity at the deduced amino acid level (Vothknecht *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 93:9287-9291 (1996), the entirety of which is herein incorporated by reference). Two different GluTR genes have been isolated from three higher plants; *Arabidopsis* (Ilag *et al.*, *Plant Cell* 6:265-75 (1994)), barley (Bougri and Grimm, *Plant J.* 9:867-878 (1996), the entirety of which is herein incorporated by reference), and cucumber (Masuda *et al.*, *Plant Cell Physiol.* 36:1237-43 (1995), the entirety of which is herein incorporated by reference). In *Arabidopsis* and cucumber, one GluTR gene is expressed in all tissues and a second is expressed in a tissue specific manner. These genes are also reported to be differentially regulated by light (Tanaka *et al.*, *Plant Physiol.* 110:1223-30 (1996); Masuda *et al.*, *Plant Physiol. Biochem.* 34:11-16 (1996); Ilag *et al.*, *Plant Cell* 6:265-75 (1994); Masuda *et al.*, *Plant Cell Physiol.* 36:1237-43 (1995); Kumar *et al.*, *Plant Mol. Biol.* 30:419-26 (1996); Hori *et al.*, *Plant Physiol. Biochem.* 34:3-9 (1996), the entirety of which is herein incorporated by reference).

GSA-AT (glutamate-1-semialdehyde aminotransferase (E.C. 5.4.3.8)), catalyzes the conversion of GSA to ALA. GSA-AT is a soluble protein localized in the chloroplast and encoded in the nucleus (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). It has a subunit molecular weight of about 45 kD. The holoenzyme consists of two identical subunits and utilizes pyridoxal phosphate ("PLP") as a cofactor (Kumar *et al.*, *Trends in Plant Science* 1:371-376 (1996); Gough *et al.*, *Glutamate 1-semialdehyde aminotransferase as a target for herbicides*, Boeger, Ed., Lewis, Boca Raton,

Fl., (1993), the entirety of which is herein incorporated by reference). GSA-AT is reported to be inhibited by gabaculine, which has also been shown to inhibit chlorophyll biosynthesis in barley leaves (Rogers and Smith, *BCPC Monogr.* 42:183-93 (1989), the entirety of which is herein incorporated by reference). GSA-AT has been crystallized from *Synechococcus* (Hennig *et al.*, *J. Mol. Biol.* 242:591-594 (1994); Hennig *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 94:4866-4871 (1997), both of which are herein incorporated by reference in their entirety).

GSA-AT genes have been cloned from a number of plants including *Arabidopsis*. The deduced amino acid sequences from plants are highly conserved. As with GluTR, two GSA-AT genes have been found in *Arabidopsis* and they may be differentially regulated by light. It has been reported that the presence of two genes for both enzymes of the C5 pathway indicate that there are two routes for ALA formation in chloroplasts (Kumar *et al.*, *Trends in Plant Science* 1:371-376 (1996)). Transgenic tobacco plants that express antisense RNA to GSA-AT have been reported to show varying degrees of chlorophyll deficiency. Antisense plants with chlorophyll contents less than about 25% of that in the wild type plants which were maintained in the greenhouse under high light conditions, did not survive (Hennig *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 94:4866-4871 (1997); Hoefgen, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 91:1726-1730 (1994), both of which are herein incorporated by reference in their entirety).

The second phase of the biosynthesis of tetrapyrroles involves the formation of the porphyrin ring. The intermediates involved in this portion of the chlorophyll/heme biosynthetic pathway, from ALA to protoporphyrin IX, appear to be essentially the same in all organisms including plants and mammals.

Porphobilinogen synthase (E.C. 4.2.1.24), also known as ALA dehydratase, catalyzes the asymmetric condensation of two molecules of ALA to yield porphobilinogen (Porra,

*Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). Porphobilinogen synthase is a metalloenzyme and there are different types of the enzyme categorized according to metal ion usage. Porphobilinogen synthase has been identified in several plants including spinach, pea, tomato, radish, and soybean. In higher plants the enzyme is located in the plastid, is a hexamer (40-50 kD subunits) and binds  $Mg^{+2}$ . The mammalian enzyme is an octamer and binds  $Zn^{2+}$  (Cheung *et al.*, *Biochemistry* 36:1148-1156 (1997); Senior *et al.*, *Biochem. J.* 320:401-412 (1996), both of which are herein incorporated by reference in their entirety). Several studies have shown that porphobilinogen synthase is both developmentally and light regulated in plants (Kyriacou *et al.*, *J. Am. Soc. Hortic. Sci.* 121:91-95 (1996), the entirety of which is herein incorporated by reference in its entirety).

Hydroxymethylbilane synthase (E.C. 4.3.1.8), also known as porphobilinogen deaminase, catalyzes the formation of the linear tetrapyrrole hydroxymethylbilane (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). The reaction involves the deamination and polymerization of four molecules of the monopyrrole porphobilinogen. Hydroxymethylbilane synthase is unusual in that it contains a novel dipyrromethane cofactor at the active site, which is self-assembled by the apoenzyme and is covalently attached to an invariant cysteine. The enzyme has been identified in mammals, yeast, bacteria, and plants (*e.g.*, pea, spinach, *Arabidopsis*). Hydroxymethylbilane synthase exists as a monomer with a molecular weight of 33-44 kD. Hydroxymethylbilane synthase from *Arabidopsis* has been cloned and found to be localized in the plastid in both roots and leaves (Witty *et al.*, *Planta* 199:557-564 (1996), the entirety of which is herein incorporated by reference). The 3-dimensional structure of porphobilinogen deaminase from *E. coli* has been

determined (Louie *et al.*, *Proteins: Struct., Funct., Genet.* 25:48-78 (1996), the entirety of which is herein incorporated by reference).

Uroporphyrinogen III (co)synthase (E.C. 4.2.1.75) catalyzes the ring closure of the unstable linear tetrapyrrole hydroxymethylbilane and the simultaneous isomerization of the acetyl and propionyl groups at pyrrole ring D forming uroporphyrinogen III (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). Uroporphyrinogen III (co)synthase has been isolated from a number of sources including mammals, bacteria, and plants (spinach). Uroporphyrinogen III (co)synthase has a molecular weight of about 30 kD and is highly diverse in primary structure depending on the source.

Uroporphyrinogen III decarboxylase (E.C. 4.1.1.37) catalyzes the stepwise decarboxylation of all four acetate side chains of uroporphyrinogen III starting with ring D followed by rings A, B, and C, respectively, to form coproporphyrinogen III (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). At high substrate concentrations, decarboxylation can occur randomly. Uroporphyrinogen III decarboxylase has been isolated from mammals, yeast, bacteria and plants (*e.g.*, tobacco, barley). It is a monomeric enzyme with a molecular weight of about 40 kD. The barley and tobacco enzymes are reported to be light regulated (Mock *et al.*, *Plant Mol. Biol.* 28:245-256 (1995), the entirety of which is herein incorporated by reference). Antisense tobacco plants have been generated and decreased levels of the enzyme were accompanied by a light-dependent necrotic phenotype and accumulation of uroporphyrinogen. It has been reported that the lesions may be caused by reactive oxygen species generated by photooxidized

uroporphyrinogen (Mock *et al.*, *Plant Mol. Biol.* 28:245-256 (1995), the entirety of which is herein incorporated by reference).

In aerobic organisms including plants, coproporphyrinogen III oxidase (E.C. 1.3.3.3), catalyzes the oxygen dependent sequential oxidative decarboxylation of the A and B propionyl side chains of coproporphyrinogen III to yield two vinyl groups and protoporphyrinogen IX (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). A separate enzyme is reported to catalyze the anaerobic reaction.

Coproporphyrinogen III oxidase has been studied in a number of organisms including plants (tobacco, pea). The enzyme is a homodimer and has a subunit molecular weight of about 35-40 kD and is located in plastids. It has been reported that coproporphyrinogen III oxidase is peripherally associated with the membrane. It has been isolated from soybean, barley and tobacco and these sequences show 70% identity at the amino acid level. Transcript levels are reportedly similar in etiolated and green leaves (barley) but higher in developing cells than in mature cells (Kruse *et al.*, *Planta* 196:796-803 (1995), the entirety of which is herein incorporated by reference). Antisense tobacco plants have been reported with decreased levels of the enzyme. The decreased level was accompanied by accumulation of coproporphyrinogen, slightly reduced chlorophyll content and a necrotic phenotype. The prominent phenotype indicates photodynamic damage (Kruse *et al.*, *EMBO J.* 14:3712-3720 (1995), the entirety of which is herein incorporated by reference).

Protoporphyrinogen IX oxidase (E.C. 1.3.3.4) catalyzes the formation of the aromatic protoporphyrin IX by the six electron oxidation of protoporphyrinogen IX (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). This is the last reported common step in tetrapyrrole biosynthesis. In aerobic

organisms, the reaction is catalyzed by a flavoprotein that utilizes oxygen as an oxidant and, under anaerobic conditions, the oxidation is achieved by passing electrons to the electron transport chain. The enzyme has been purified from a number of sources including mammals and plants (barley) and is an integral membrane protein. The barley enzyme has a molecular weight of 36 kD and activity has been found in both plastidal and mitochondrial extracts.

The plastidal and mitochondrial forms of protoporphyrinogen IX oxidase have been cloned from tobacco and were found to exhibit low homology. The mitochondrial form is associated with heme biosynthesis. The plastidic enzyme functions primarily in the formation of chlorophyll and to a lesser extent in the formation of heme required for plastid proteins (Lermontova *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 94:8895-8900 (1997), the entirety of which is herein incorporated by reference). Protoporphyrinogen IX oxidase is susceptible to inhibition by a number of herbicides including diphenyl ethers. Phytotoxicity has been explained as due to the accumulation of excess protoporphyrinogen which is rapidly oxidized to protoporphyrin in the cytoplasm. Protoporphyrin has been reported as a potent photosensitizer which generates singlet oxygen and causes rapid lipid peroxidation and cell death.

In the third and final phase of tetrapyrrole biosynthesis, magnesium or iron is inserted into protoporphyrin IX and subsequent modifications lead to the synthesis of the final tetrapyrrole products, such as chlorophyll and heme.

Mg-chelatase catalyzes the conversion of protoporphyrin IX to magnesium protoporphyrin IX by the insertion of  $Mg^{+2}$  (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). Mg-chelatase, which requires ATP, is reportedly a three component enzyme. The three protein components have molecular weights of about 140, 40, and 70 kD. The reaction takes place in two steps, an ATP-dependent

activation followed by an ATP-dependent chelation step. Mg-chelatase activity has been demonstrated in peas, cucumber, and barley and reportedly is localized in the chloroplast. Barley, *Arabidopsis*, and soybean genes encoding the 140 and 40 kD subunits have been cloned. Studies with the two identified plant genes show that Mg-chelatase expression is light regulated (Walker and Willows, *Biochem. J.* 327:321-333 (1997), the entirety of which is herein incorporated by reference).

Mg-protoporphyrin IX *O*-methyltransferase (E.C. 2.1.1.11) esterifies the propionic side chain of ring III of Mg-protoporphyrin IX to form Mg-protoporphyrin IX monomethylester (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). The methyl group is donated by the cofactor S-adenosyl-L-methionine. The enzyme has been isolated from bacteria and plants (wheat). The gene for Mg-protoporphyrin IX *O*-methyltransferase has been cloned from bacteria including *Synechocystis* (Smith *et al.*, *Plant Mol. Biol.* 30:1307-1314 (1996), the entirety of which is herein incorporated by reference).

Mg-protoporphyrin IX monomethyl ester cyclase catalyzes the cyclization of Mg-protoporphyrin IX monomethylester to form the isocyclic ring E of divinyl protochlorophyllide (Porra, *Photochemistry and Photobiology* 65:492-516 (1997)). In aerobic organisms the enzymatic reaction is dependent on O<sub>2</sub> and NADPH. Evidence suggests that Mg-protoporphyrin IX monomethyl ester cyclase is a membrane-bound monooxygenase of the iron-sulfur protein or copper protein type. Mg-protoporphyrin IX monomethyl ester cyclase has been extracted from chloroplasts of higher plants including cucumber and wheat. A cucumber enzyme has been shown to consist of two components, a soluble and a membrane-bound component. The soluble

component has a molecular weight of 30 kD (Bollivar and Beale, *Plant Physiol.* 112:105-114 (1996), the entirety of which is herein incorporated by reference).

The reduction of divinyl protochlorophyllide to monovinyl protochlorophyllide has been reported based on product characterization, this reaction is catalyzed by 8-vinyl reductase (Porra, *Photochemistry and Photobiology* 65:492-516 (1997)). It has been reported that Mg-protoporphyrin IX monomethylester may also act as a substrate. NADPH is the most likely reductant. 8-vinyl reductase has been detected in higher plants including wheat and cucumber.

Protochlorophyllide reductase ("POR") (E.C. 1.3.1.33) catalyzes the reduction of the double bond between carbons 7 and 8 of the D ring of protochlorophyllide producing chlorophyllide (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). In angiosperms this is a light-dependent reaction. Non-flowering land plants, algae, and cyanobacteria contain both a light-dependent and a light-independent enzyme. Some other organisms contain only the light-independent enzyme. Three chloroplast genes have been identified that are essential for the light-independent enzyme (chlL, chlN and chlB).

The light-dependent POR ("L-POR") has been purified from barley, oat, and *Arabidopsis*. L-POR has a molecular weight of 35-38 kD and forms different multimers and aggregates with other proteins. L-POR is localized in the plastid and encoded in the nucleus. The genes encoding L-POR have been cloned from, for example, barley, *Arabidopsis*, pea, and oat. Two distinct and differentially light-regulated L-POR genes, POR A and POR B, have been identified in *Arabidopsis* and barley. POR A and POR B have biochemically equivalent light-dependent activities, with different expression patterns. POR B is reported to be present throughout the plant life cycle, while POR A is reported to function only in the very early stages of greening of



etiolated tissue (Runge *et al.*, *Plant J.* 9:513-523 (1996); Holtorf and Apel, *Plant Mol. Biol.* 31:387-392 (1996); Martin *et al.*, *Biochem. J.* 325:139-145 (1997), all of which are herein incorporated by reference in their entirety).

Chlorophyll synthetase catalyzes the last reported step in chlorophyll *a* biosynthesis (Porra, *Photochemistry and Photobiology* 65:492-516 (1997); von Wettstein *et al.*, *Plant Cell* 7:1039-1057 (1995)). Chlorophyll synthetase esterifies the propionic acid side chain of ring D of chlorophyllide with either phytyl pyrophosphate in green plants or geranylgeranyl pyrophosphate in greening etiolated seedlings. The enzyme is located in the plastid. A gene that encodes the enzyme in *Synechocystis* (chlG) and a gene that encodes the enzyme in *Arabidopsis* (G4) have been cloned and expressed in *E. coli*. The *Synechocystis* enzyme has the preferred substrate specificity reported for green plants. The cloned and expressed enzyme from *Arabidopsis* has the preferred substrate specificity reported for etiolated plants (Oster *et al.*, *J. Biol. Chem.* 272:9671-9676 (1997); Oster and Rudiger, *Bot. Acta* 110:420-423 (1997), both of which are herein incorporated by reference).

Ferrochelatase (E.C. 4.99.1.1) catalyzes the conversion of protoporphyrin IX to heme. In plants the enzyme is located in both mitochondria and plastids. Ferrochelatase is reported to be a single soluble protein. Two ferrochelatase genes have been identified in *Arabidopsis*. Ferrochelatase-II encodes a protein targeted to the chloroplast and ferrochelatase-I encodes a protein targeted to both chloroplasts and mitochondria (Roper and Smith, *Eur. J. Biochem.* 246:32-37 (1997); Chow *et al.*, *J. Biol. Chem.* 272:27565-27571 (1997), both of which are herein incorporated by reference).

## II. EXPRESSED SEQUENCE TAG NUCLEIC ACID MOLECULES

Expressed sequence tags, or ESTs are randomly sequenced members of a cDNA library (or complementary DNA)(McCombie *et al.*, *Nature Genetics* 1:124-130 (1992); Kurata *et al.*, *Nature Genetics* 8:365-372 (1994); Okubo *et al.*, *Nature Genetics* 2:173-179 (1992), all of which references are incorporated herein in their entirety). The randomly selected clones comprise insets that can represent a copy of up to the full length of a mRNA transcript.

Using conventional methodologies, cDNA libraries can be constructed from the mRNA (messenger RNA) of a given tissue or organism using poly dT primers and reverse transcriptase (Efstratiadis *et al.*, *Cell* 7:279-3680 (1976), the entirety of which is herein incorporated by reference; Higuchi *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 73:3146-3150 (1976), the entirety of which is herein incorporated by reference; Maniatis *et al.*, *Cell* 8:163-182 (1976) the entirety of which is herein incorporated by reference; Land *et al.*, *Nucleic Acids Res.* 9:2251-2266 (1981), the entirety of which is herein incorporated by reference; Okayama *et al.*, *Mol. Cell. Biol.* 2:161-170 (1982), the entirety of which is herein incorporated by reference; Gubler *et al.*, *Gene* 25:263-269 (1983), the entirety of which is herein incorporated by reference).

Several methods may be employed to obtain full-length cDNA constructs. For example, terminal transferase can be used to add homopolymeric tails of dC residues to the free 3' hydroxyl groups (Land *et al.*, *Nucleic Acids Res.* 9:2251-2266 (1981), the entirety of which is herein incorporated by reference). This tail can then be hybridized by a poly dG oligo which can act as a primer for the synthesis of full length second strand cDNA. Okayama and Berg, *Mol. Cell. Biol.* 2:161-170 (1982), the entirety of which is herein incorporated by reference, report a method for obtaining full length cDNA constructs. This method has been simplified by using synthetic primer-adapters that have both homopolymeric tails for priming the synthesis of the

first and second strands and restriction sites for cloning into plasmids (Coleclough *et al.*, *Gene* 34:305-314 (1985), the entirety of which is herein incorporated by reference) and bacteriophage vectors (Krawinkel *et al.*, *Nucleic Acids Res.* 14:1913 (1986), the entirety of which is herein incorporated by reference; Han *et al.*, *Nucleic Acids Res.* 15:6304 (1987), the entirety of which is herein incorporated by reference).

These strategies have been coupled with additional strategies for isolating rare mRNA populations. For example, a typical mammalian cell contains between 10,000 and 30,000 different mRNA sequences (Davidson, *Gene Activity in Early Development*, 2nd ed., Academic Press, New York (1976), the entirety of which is herein incorporated by reference). The number of clones required to achieve a given probability that a low-abundance mRNA will be present in a cDNA library is  $N = (\ln(1-P))/(\ln(1-1/n))$  where N is the number of clones required, P is the probability desired and 1/n is the fractional proportion of the total mRNA that is represented by a single rare mRNA (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press (1989), the entirety of which is herein incorporated by reference).

A method to enrich preparations of mRNA for sequences of interest is to fractionate by size. One such method is to fractionate by electrophoresis through an agarose gel (Pennica *et al.*, *Nature* 301:214-221 (1983), the entirety of which is herein incorporated by reference). Another such method employs sucrose gradient centrifugation in the presence of an agent, such as methylmercuric hydroxide, that denatures secondary structure in RNA (Schweinfest *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 79:4997-5000 (1982), the entirety of which is herein incorporated by reference).

A frequently adopted method is to construct equalized or normalized cDNA libraries (Ko, *Nucleic Acids Res.* 18:5705-5711 (1990), the entirety of which is herein incorporated by reference; Patanjali *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 88:1943-1947 (1991), the entirety of which is herein incorporated by reference). Typically, the cDNA population is normalized by subtractive hybridization (Schmid *et al.*, *J. Neurochem.* 48:307-312 (1987), the entirety of which is herein incorporated by reference; Fargnoli *et al.*, *Anal. Biochem.* 187:364-373 (1990), the entirety of which is herein incorporated by reference; Travis *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 85:1696-1700 (1988), the entirety of which is herein incorporated by reference; Kato, *Eur. J. Neurosci.* 2:704-711 (1990); and Schweinfest *et al.*, *Genet. Anal. Tech. Appl.* 7:64-70 (1990), the entirety of which is herein incorporated by reference). Subtraction represents another method for reducing the population of certain sequences in the cDNA library (Swaroop *et al.*, *Nucleic Acids Res.* 19:1954 (1991), the entirety of which is herein incorporated by reference).

ESTs can be sequenced by a number of methods. Two basic methods may be used for DNA sequencing, the chain termination method of Sanger *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 74:5463-5467 (1977), the entirety of which is herein incorporated by reference and the chemical degradation method of Maxam and Gilbert, *Proc. Nat. Acad. Sci. (U.S.A.)* 74:560-564 (1977), the entirety of which is herein incorporated by reference. Automation and advances in technology such as the replacement of radioisotopes with fluorescence-based sequencing have reduced the effort required to sequence DNA (Craxton, *Methods* 2:20-26 (1991), the entirety of which is herein incorporated by reference; Ju *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 92:4347-4351 (1995), the entirety of which is herein incorporated by reference; Tabor and Richardson, *Proc. Natl. Acad. Sci. (U.S.A.)* 92:6339-6343 (1995), the entirety of which is herein incorporated by reference). Automated sequencers are available from, for example, Pharmacia Biotech, Inc.,

Piscataway, New Jersey (Pharmacia ALF), LI-COR, Inc., Lincoln, Nebraska (LI-COR 4,000) and Millipore, Bedford, Massachusetts (Millipore BaseStation).

In addition, advances in capillary gel electrophoresis have also reduced the effort required to sequence DNA and such advances provide a rapid high resolution approach for sequencing DNA samples (Swerdlow and Gesteland, *Nucleic Acids Res.* 18:1415-1419 (1990); Smith, *Nature* 349:812-813 (1991); Luckey *et al.*, *Methods Enzymol.* 218:154-172 (1993); Lu *et al.*, *J. Chromatog. A.* 680:497-501 (1994); Carson *et al.*, *Anal. Chem.* 65:3219-3226 (1993); Huang *et al.*, *Anal. Chem.* 64:2149-2154 (1992); Kheterpal *et al.*, *Electrophoresis* 17:1852-1859 (1996); Quesada and Zhang, *Electrophoresis* 17:1841-1851 (1996); Baba, *Yakugaku Zasshi* 117:265-281 (1997), all of which are herein incorporated by reference in their entirety).

ESTs longer than 150 nucleotides have been found to be useful for similarity searches and mapping (Adams *et al.*, *Science* 252:1651-1656 (1991), herein incorporated by reference). ESTs, which can represent copies of up to the full length transcript, may be partially or completely sequenced. Between 150-450 nucleotides of sequence information is usually generated as this is the length of sequence information that is routinely and reliably produced using single run sequence data. Typically, only single run sequence data is obtained from the cDNA library (Adams *et al.*, *Science* 252:1651-1656 (1991). Automated single run sequencing typically results in an approximately 2-3% error or base ambiguity rate (Boguski *et al.*, *Nature Genetics* 4:332-333 (1993), the entirety of which is herein incorporated by reference).

EST databases have been constructed or partially constructed from, for example, *C. elegans* (McCombie *et al.*, *Nature Genetics* 1:124-131 (1992)), human liver cell line HepG2 (Okubo *et al.*, *Nature Genetics* 2:173-179 (1992)), human brain RNA (Adams *et al.*, *Science* 252:1651-1656 (1991); Adams *et al.*, *Nature* 355:632-635 (1992)), *Arabidopsis*, (Newman *et al.*,

*Plant Physiol.* 106:1241-1255 (1994)); and rice (Kurata *et al.*, *Nature Genetics* 8:365-372 (1994)).

### III. SEQUENCE COMPARISONS

A characteristic feature of a DNA sequence is that it can be compared with other DNA sequences. Sequence comparisons can be undertaken by determining the similarity of the test or query sequence with sequences in publicly available or proprietary databases (“similarity analysis”) or by searching for certain motifs (“intrinsic sequence analysis”)(e.g. *cis* elements)(Coulson, *Trends in Biotechnology* 12:76-80 (1994), the entirety of which is herein incorporated by reference); Birren *et al.*, *Genome Analysis 1*: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 543-559 (1997), the entirety of which is herein incorporated by reference).

Similarity analysis includes database search and alignment. Examples of public databases include the DNA Database of Japan (DDBJ)(<http://www.ddbj.nig.ac.jp/>); Genbank (<http://www.ncbi.nlm.nih.gov/Web/Search/Index.html>); and the European Molecular Biology Laboratory Nucleic Acid Sequence Database (EMBL) ([http://www.ebi.ac.uk/ebi\\_docs/embl\\_db/embl-db.html](http://www.ebi.ac.uk/ebi_docs/embl_db/embl-db.html)). Other appropriate databases include dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>), SwissProt ([http://www.ebi.ac.uk/ebi\\_docs/swisprot\\_db/swisshome.html](http://www.ebi.ac.uk/ebi_docs/swisprot_db/swisshome.html)), PIR (<http://www-nbrt.georgetown.edu/pir/>) and The Institute for Genome Research (<http://www.tigr.org/tdb/tdb.html>)

A number of different search algorithms have been developed, one example of which are the suite of programs referred to as BLAST programs. There are five implementations of BLAST, three designed for nucleotide sequences queries (BLASTN, BLASTX and TBLASTX)

and two designed for protein sequence queries (BLASTP and TBLASTN) (Coulson, *Trends in Biotechnology* 12:76-80 (1994); Birren *et al.*, *Genome Analysis 1*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 543-559 (1997)).

BLASTN takes a nucleotide sequence (the query sequence) and its reverse complement and searches them against a nucleotide sequence database. BLASTN was designed for speed, not maximum sensitivity and may not find distantly related coding sequences. BLASTX takes a nucleotide sequence, translates it in three forward reading frames and three reverse complement reading frames and then compares the six translations against a protein sequence database. BLASTX is useful for sensitive analysis of preliminary (single-pass) sequence data and is tolerant of sequencing errors (Gish and States, *Nature Genetics* 3:266-272 (1993), the entirety of which is herein incorporated by reference). BLASTN and BLASTX may be used in concert for analyzing EST data (Coulson, *Trends in Biotechnology* 12:76-80 (1994); Birren *et al.*, *Genome Analysis 1*:543-559 (1997)).

Given a coding nucleotide sequence and the protein it encodes, it is often preferable to use the protein as the query sequence to search a database because of the greatly increased sensitivity to detect more subtle relationships. This is due to the larger alphabet of proteins (20 amino acids) compared with the alphabet of nucleic acid sequences (4 bases), where it is far easier to obtain a match by chance. In addition, with nucleotide alignments, only a match (positive score) or a mismatch (negative score) is obtained, but with proteins, the presence of conservative amino acid substitutions can be taken into account. Here, a mismatch may yield a positive score if the non-identical residue has physical/chemical properties similar to the one it replaced. Various scoring matrices are used to supply the substitution scores of all possible amino acid pairs. A general purpose scoring system is the BLOSUM62 matrix (Henikoff and

Henikoff, *Proteins* 17:49-61 (1993), the entirety of which is herein incorporated by reference), which is currently the default choice for BLAST programs. BLOSUM62 is tailored for alignments of moderately diverged sequences and thus may not yield the best results under all conditions. Altschul, *J. Mol. Biol.* 36:290-300 (1993), the entirety of which is herein incorporated by reference, describes a combination of three matrices to cover all contingencies. This may improve sensitivity, but at the expense of slower searches. In practice, a single BLOSUM62 matrix is often used but others (PAM40 and PAM250) may be attempted when additional analysis is necessary. Low PAM matrices are directed at detecting very strong but localized sequence similarities, whereas high PAM matrices are directed at detecting long but weak alignments between very distantly related sequences.

Homologues in other organisms are available that can be used for comparative sequence analysis. Multiple alignments are performed to study similarities and differences in a group of related sequences. CLUSTAL W is a multiple sequence alignment package that performs progressive multiple sequence alignments based on the method of Feng and Doolittle, *J. Mol. Evol.* 25:351-360 (1987), the entirety of which is herein incorporated by reference. Each pair of sequences is aligned and the distance between each pair is calculated; from this distance matrix, a guide tree is calculated and all of the sequences are progressively aligned based on this tree. A feature of the program is its sensitivity to the effect of gaps on the alignment; gap penalties are varied to encourage the insertion of gaps in probable loop regions instead of in the middle of structured regions. Users can specify gap penalties, choose between a number of scoring matrices, or supply their own scoring matrix for both pairwise alignments and multiple alignments. CLUSTAL W for UNIX and VMS systems is available at: <ftp.ebi.ac.uk>. Another program is MACAW (Schuler *et al.*, *Proteins Struct. Func. Genet.* 9:180-190 (1991), the entirety



of which is herein incorporated by reference, for which both Macintosh and Microsoft Windows versions are available. MACAW uses a graphical interface, provides a choice of several alignment algorithms and is available by anonymous ftp at: [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov/directory/pub/macaw) (directory/pub/macaw).

Sequence motifs are derived from multiple alignments and can be used to examine individual sequences or an entire database for subtle patterns. With motifs, it is sometimes possible to detect distant relationships that may not be demonstrable based on comparisons of primary sequences alone. Currently, the largest collection of sequence motifs in the world is PROSITE (Bairoch and Bucher, *Nucleic Acid Research* 22:3583-3589 (1994), the entirety of which is herein incorporated by reference). PROSITE may be accessed via either the ExPASy server on the World Wide Web or anonymous ftp site. Many commercial sequence analysis packages also provide search programs that use PROSITE data.

A resource for searching protein motifs is the BLOCKS E-mail server developed by Henikoff, *Trends Biochem Sci.* 18:267-268 (1993), the entirety of which is herein incorporated by reference; Henikoff and Henikoff, *Nucleic Acid Research* 19:6565-6572 (1991), the entirety of which is herein incorporated by reference; Henikoff and Henikoff, *Proteins* 17:49-61 (1993). BLOCKS searches a protein or nucleotide sequence against a database of protein motifs or "blocks." Blocks are defined as short, ungapped multiple alignments that represent highly conserved protein patterns. The blocks themselves are derived from entries in PROSITE as well as other sources. Either a protein query or a nucleotide query can be submitted to the BLOCKS server; if a nucleotide sequence is submitted, the sequence is translated in all six reading frames and motifs are sought for these conceptual translations. Once the search is completed, the server

will return a ranked list of significant matches, along with an alignment of the query sequence to the matched BLOCKS entries.

Conserved protein domains can be represented by two-dimensional matrices, which measure either the frequency or probability of the occurrences of each amino acid residue and deletions or insertions in each position of the domain. This type of model, when used to search against protein databases, is sensitive and usually yields more accurate results than simple motif searches. Two popular implementations of this approach are profile searches such as GCG program ProfileSearch and Hidden Markov Models (HMMs)(Krough *et al.*, *J. Mol. Biol.* 235:1501-1531, (1994); Eddy, *Current Opinion in Structural Biology* 6:361-365, (1996), both of which are herein incorporated by reference in their entirety). In both cases, a large number of common protein domains have been converted into profiles, as present in the PROSITE library, or HMM models, as in the Pfam protein domain library (Sonnhammer *et al.*, *Proteins* 28:405-420 (1997), the entirety of which is herein incorporated by reference). Pfam contains more than 500 HMM models for enzymes, transcription factors, signal transduction molecules and structural proteins. Protein databases can be queried with these profiles or HMM models, which will identify proteins containing the domain of interest. For example, HMMSW or HMMFS, two programs in a public domain package called HMMER (Sonnhammer *et al.*, *Proteins* 28:405-420 (1997)) can be used.

PROSITE and BLOCKS represent collected families of protein motifs. Thus, searching these databases entails submitting a single sequence to determine whether or not that sequence is similar to the members of an established family. Programs working in the opposite direction compare a collection of sequences with individual entries in the protein databases. An example of such a program is the Motif Search Tool, or MoST (Tatusov *et al.*, *Proc. Natl. Acad. Sci.*

(U.S.A.) 91:12091-12095 (1994), the entirety of which is herein incorporated by reference). On the basis of an aligned set of input sequences, a weight matrix is calculated by using one of four methods (selected by the user). A weight matrix is simply a representation, position by position of how likely a particular amino acid will appear. The calculated weight matrix is then used to search the databases. To increase sensitivity, newly found sequences are added to the original data set, the weight matrix is recalculated and the search is performed again. This procedure continues until no new sequences are found.

### **SUMMARY OF THE INVENTION**

The present invention provides a substantially purified nucleic acid molecule that encodes a maize or soybean tetrapyrrole pathway protein or fragment thereof, wherein the maize or soybean tetrapyrrole pathway protein is selected from the group consisting of: (a) putative chlorophyll synthetase enzyme; (b) protochlorophyllide reductase enzyme; (c) putative protochlorophyllide reductase enzyme; (d) coproporphyrinogen oxidase enzyme; (e) protoporphyrinogen oxidase enzyme; (f) uroporphyrinogen decarboxylase enzyme; (g) putative uroporphyrinogen decarboxylase enzyme (h) porphobilinogen synthase enzyme; (i) hydroxymethylbilane synthase enzyme; (j) glutamate-1-semialdehyde 2,1-aminomutase enzyme; (k) glutamate tRNA ligase enzyme; (l) glutamyl-tRNA reductase enzyme; (m) Mg-chelatase enzyme, and (n) ferrochelatase enzyme.

The present invention also provides a substantially purified nucleic acid molecule that encodes a plant tetrapyrrole pathway protein or fragment thereof, wherein the nucleic acid molecule is selected from the group consisting of a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or fragment thereof, a

nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or fragment thereof and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or fragment thereof

The present invention also provides a substantially purified maize or soybean tetrapyrrole pathway protein or fragment thereof, wherein the maize or soybean tetrapyrrole pathway protein is selected from the group consisting of (a) putative chlorophyll synthetase enzyme or fragment thereof; (b) putative protochlorophyllide reductase enzyme or fragment thereof; (c) protochlorophyllide reductase enzyme or fragment thereof; (d) coproporphyrinogen oxidase enzyme or fragment thereof; (e) protoporphyrinogen oxidase enzyme or fragment thereof; (f) uroporphyrinogen decarboxylase enzyme or fragment thereof; (g) putative uroporphyrinogen decarboxylase enzyme or fragment thereof; (h) porphobilinogen synthase enzyme or fragment

thereof; (i) hydroxymethylbilane synthase enzyme or fragment thereof; (j) glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof; (k) glutamate tRNA ligase enzyme or fragment thereof; (l) glutamyl-tRNA reductase enzyme or fragment thereof; (m) Mg-chelatase enzyme or fragment thereof; and (n) ferrochelatase enzyme or fragment thereof.

The present invention also provides a substantially purified maize or soybean tetrapyrrole pathway protein or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 1 through SEQ ID NO: 677.

The present invention also provides a substantially purified maize or soybean putative chlorophyll synthetase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 1 through SEQ ID NO: 8 and SEQ ID NO: 384 through SEQ ID NO: 397.

The present invention also provides a substantially purified maize or soybean putative chlorophyll synthetase enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 8 and SEQ ID NO: 384 through SEQ ID NO: 397.

The present invention also provides a substantially purified maize or soybean protochlorophyllide reductase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a

complement SEQ ID NO: 9 through SEQ ID NO: 94 and SEQ ID NO: 398 through SEQ ID NO: 466.

The present invention also provides a substantially purified maize or soybean protochlorophyllide reductase enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 9 through SEQ ID NO: 94 and SEQ ID NO: 398 through SEQ ID NO: 466.

The present invention also provides a substantially purified maize or soybean putative protochlorophyllide reductase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement SEQ ID NO: 95 through SEQ ID NO: 96 and SEQ ID NO: 467 through SEQ ID NO: 479.

The present invention also provides a substantially purified maize or soybean putative protochlorophyllide reductase enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 95 through SEQ ID NO: 96 and SEQ ID NO: 467 through SEQ ID NO: 479.

The present invention also provides a substantially purified maize or soybean coproporphyrinogen oxidase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence consisting of a complement of SEQ ID NO: 97 through SEQ ID NO: 128 and SEQ ID NO: 480 through SEQ ID NO: 494.

The present invention also provides a substantially purified maize or soybean coproporphyrinogen oxidase enzyme or fragment thereof encoded by a nucleic acid sequence

consisting of SEQ ID NO: 97 through SEQ ID NO: 128 and SEQ ID NO: 480 through SEQ ID NO: 494.

The present invention also provides a substantially purified maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 129 through SEQ ID NO: 131 and SEQ ID NO: 495 through SEQ ID NO: 499.

The present invention also provides a substantially purified maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 129 through SEQ ID NO: 131 and SEQ ID NO: 495 through SEQ ID NO: 499.

The present invention also provides a substantially purified maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 132 through SEQ ID NO: 144 and SEQ ID NO: 500 through SEQ ID NO: 509.

The present invention also provides a substantially purified maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 132 through SEQ ID NO: 144 and SEQ ID NO: 500 through SEQ ID NO: 509.

The present invention also provides a substantially purified a maize putative uroporphyrinogen decarboxylase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence consisting of a complement of SEQ ID NO: 510.

The present invention also provides a substantially purified maize putative uroporphyrinogen decarboxylase enzyme or fragment thereof encoded by a nucleic acid sequence consisting of SEQ ID NO: 510.

The present invention also provides a substantially purified soybean porphobilinogen synthetase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 145 through SEQ ID NO: 191 and SEQ ID NO: 511 through SEQ ID NO: 531.

The present invention also provides a substantially purified maize or soybean porphobilinogen synthetase enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 145 through SEQ ID NO: 191 and SEQ ID NO: 511 through SEQ ID NO: 531.

The present invention also provides a substantially purified maize or soybean hydroxymethylbilane synthase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement SEQ ID NO: 154, SEQ ID NO: 192 through SEQ ID NO: 217 and SEQ ID NO: 532 through SEQ ID NO: 542.



The present invention also provides a substantially purified maize or soybean hydroxymethylbilane enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 154, SEQ ID NO: 192 through SEQ ID NO: 217 and SEQ ID NO: 532 through SEQ ID NO: 542.

The present invention also provides a substantially purified maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence consisting of a complement of SEQ ID NO: 218 through SEQ ID NO: 265 and SEQ ID NO: 543 through SEQ ID NO: 569.

The present invention also provides a substantially purified maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof encoded by a nucleic acid sequence consisting of SEQ ID NO: 218 through SEQ ID NO: 265 and SEQ ID NO: 543 through SEQ ID NO: 569.

The present invention also provides a substantially purified maize or soybean glutamate tRNA ligase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 266 through SEQ ID NO: 289 and SEQ ID NO: 570 through SEQ ID NO: 585.

The present invention also provides a substantially purified maize or soybean glutamate tRNA ligase enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 266 through SEQ ID NO: 289 and SEQ ID NO: 570 through SEQ ID NO: 585.

560210" 872220  
SUB D

The present invention also provides a substantially purified maize or soybean glutamyl-tRNA reductase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 266 through SEQ ID NO: 289 and SEQ ID NO: 570 through SEQ ID NO: 585.

The present invention also provides a substantially purified maize or soybean glutamyl-tRNA reductase enzyme fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 266 through SEQ ID NO: 289 and SEQ ID NO: 570 through SEQ ID NO: 585.

The present invention also provides a substantially purified maize or soybean Mg-chelatase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 290 through SEQ ID NO: 306 and SEQ ID NO: 586 through SEQ ID NO: 609.

The present invention also provides a substantially purified maize or soybean Mg-chelatase enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 290 through SEQ ID NO: 306 and SEQ ID NO: 586 through SEQ ID NO: 609.

The present invention also provides a substantially purified maize or soybean ferrochelatase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 372 through SEQ ID NO: 383 and SEQ ID NO: 653 through SEQ ID NO: 677.

The present invention also provides a substantially purified maize or soybean ferrochelatase enzyme or fragment thereof encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO: 372 through SEQ ID NO: 383 and SEQ ID NO: 653 through SEQ ID NO: 677.

The present invention also provides a purified antibody or fragment thereof which is capable of specifically binding to a maize or soybean tetrapyrrole pathway protein or fragment thereof, wherein the maize or soybean tetrapyrrole pathway protein or fragment thereof is encoded by a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a putative maize or soybean chlorophyll synthetase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 1 through SEQ ID NO: 8 and SEQ ID NO: 384 through SEQ ID NO: 397 or a nucleic acid sequence selected from the group consisting SEQ ID NO: 1 through SEQ ID NO: 8 and SEQ ID NO: 384 through SEQ ID NO: 397.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean protochlorophyllide reductase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 9 through SEQ ID NO: 94 and SEQ ID NO: 398 through SEQ ID

SUB  
D2  
CDU+

NO: 466 or a nucleic acid sequence selected from the group consisting of SEQ ID NO: 95 through SEQ ID NO: 96 and SEQ ID NO: 398 through SEQ ID NO: 466.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean putative protochlorophyllide reductase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 95 through SEQ ID NO: 96 and SEQ ID NO: 467 through SEQ ID NO: 479 or a nucleic acid sequence selected from the group consisting of SEQ ID NO: 95 through SEQ ID NO: 96 and SEQ ID NO: 467 through SEQ ID NO: 479.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean coproporphyrinogen oxidase or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule consisting of a complement of a nucleic acid sequence having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 97 through SEQ ID NO: 128 and SEQ ID NO: 480 through SEQ ID NO: 494 or a nucleic acid sequence selected from the group consisting of SEQ ID NO: 97 through SEQ ID NO: 128 and SEQ ID NO: 480 through SEQ ID NO: 494.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a

complement of SEQ ID NO: 129 through SEQ ID NO: 131 and SEQ ID NO: 495 through SEQ ID NO: 499 or a nucleic acid sequence selected from the group consisting SEQ ID NO: 129 through SEQ ID NO: 131 and SEQ ID NO: 495 through SEQ ID NO: 499.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 132 through SEQ ID NO: 144 and SEQ ID NO: 500 through SEQ ID NO: 509 or a nucleic acid sequence selected from the group consisting SEQ ID NO: 132 through SEQ ID NO: 144 and SEQ ID NO: 500 through SEQ ID NO: 509.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a putative maize uroporphyrinogen decarboxylase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence consisting of a complement of SEQ ID NO: 510 or a nucleic acid sequence consisting SEQ ID NO: 510.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean porphobilinogen enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 145 through SEQ ID NO: 191 and SEQ ID NO: 511 through SEQ ID NO: 531 or a nucleic

acid sequence selected from the group consisting SEQ ID NO: 145 through SEQ ID NO: 191 and SEQ ID NO: 511 through SEQ ID NO: 531.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean hydroxymethylbilane synthase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 154, SEQ ID NO: 192 through SEQ ID NO: 217 and SEQ ID NO: 532 through SEQ ID NO: 542 or a nucleic acid sequence selected from the group consisting of SEQ ID NO: 154, SEQ ID NO: 192 through SEQ ID NO: 217 and SEQ ID NO: 532 through SEQ ID NO: 542.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule consisting of a compliment of a nucleic acid sequence having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 218 through SEQ ID NO: 265 and SEQ ID NO: 543 through SEQ ID NO: 569 or a nucleic acid sequence selected from the group consisting of SEQ ID NO: 218 through SEQ ID NO: 265 and SEQ ID NO: 543 through SEQ ID NO: 569.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean glutamate tRNA ligase enzyme or fragment thereof encoded by a first nucleic acid molecule which

specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 266 through SEQ ID NO: 289 and SEQ ID NO: 570 through SEQ ID NO: 585 or a nucleic acid sequence selected from the group consisting SEQ ID NO: 266 through SEQ ID NO: 289 and SEQ ID NO: 570 through SEQ ID NO: 585.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean glutamyl-tRNA reductase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 290 through SEQ ID NO: 306 and SEQ ID NO: 586 through SEQ ID NO: 609 or a nucleic acid sequence selected from the group consisting SEQ ID NO: 290 through SEQ ID NO: 306 and SEQ ID NO: 586 through SEQ ID NO: 609.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean Mg-chelatase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 307 through SEQ ID NO: 371 and SEQ ID NO: 610 through SEQ ID NO: 652 or a nucleic acid sequence selected from the group consisting SEQ ID NO: 307 through SEQ ID NO: 371 and SEQ ID NO: 610 through SEQ ID NO: 652.

The present invention also provides a substantially purified antibody or fragment thereof, the antibody or fragment thereof capable of specifically binding to a maize or soybean

ferrochelatase enzyme or fragment thereof encoded by a first nucleic acid molecule which specifically hybridizes to a second nucleic acid molecule, the second nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 372 through SEQ ID NO: 383 and SEQ ID NO: 653 through SEQ ID NO: 677 or a nucleic acid sequence selected from the group consisting SEQ ID NO: 372 through SEQ ID NO: 383 and SEQ ID NO: 653 through SEQ ID NO: 677.

The present invention also provides a transformed plant having a nucleic acid molecule which comprises: (A) an exogenous promoter region which functions in a plant cell to cause the production of a mRNA molecule; (B) a structural nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of (a) a nucleic acid sequence which encodes for a putative chlorophyll synthetase enzyme or fragment thereof; (b) a nucleic acid sequence which encodes for a protochlorophyllide reductase or fragment thereof; (c) a nucleic acid sequence which encodes for a putative protochlorophyllide reductase or fragment thereof; (d) a nucleic acid sequence which encodes for a coproporphyrinogen oxidase or fragment thereof; (e) a nucleic acid sequence which encodes for a protoporphyrinogen oxidase enzyme or fragment thereof; (f) a nucleic acid sequence which encodes for a uroporphyrinogen decarboxylase enzyme or fragment thereof; (g) a nucleic acid sequence which encodes for a putative uroporphyrinogen decarboxylase enzyme or fragment thereof; (h) a nucleic acid sequence which encodes for a porphobilinogen synthase enzyme or fragment thereof; (i) a nucleic acid sequence which encodes for a hydroxymethylbilane synthase enzyme or fragment thereof; (j) a nucleic acid sequence which encodes for a glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof; (k) a nucleic acid sequence which encodes for a glutamate tRNA ligase enzyme or fragment thereof; (l) a nucleic acid sequence which encodes for a glutamyl-tRNA reductase



enzyme or fragment thereof; (m) a nucleic acid sequence which encodes for a Mg-chelatase enzyme or fragment thereof; and (n) a nucleic acid sequence which encodes for a ferrochelatase enzyme or fragment thereof (m) a nucleic acid sequence which is complementary to any of the nucleic acid sequences of (a) through (n); and (C) a 3' non-translated sequence that functions in the plant cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule.

The present invention also provides a transformed plant having a nucleic acid molecule which comprises: (A) an exogenous promoter region which functions in a plant cell to cause the production of a mRNA molecule; which is linked to (B) a structural nucleic acid molecule, wherein the structural nucleic acid molecule encodes a plant tetrapyrrole pathway protein or fragment thereof, the structural nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or fragment thereof; which is linked to (C) a 3' non-translated sequence that functions in the plant cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule.

The present invention also provides a transformed plant having a nucleic acid molecule which comprises: (A) an exogenous promoter region which functions in a plant cell to cause the production of a mRNA molecule; which is linked to (B) a structural nucleic acid molecule, wherein the structural nucleic acid molecule is selected from the group consisting of a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or fragment thereof, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or fragment thereof, a nucleic acid sequence

which encodes a maize or soybean coproporphyrinogen oxidase or fragment thereof, a nucleic acid sequence which encodes a maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof, a nucleic acid sequence which encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof, a nucleic acid sequence which encodes a putative maize uroporphyrinogen decarboxylase enzyme or fragment thereof, a nucleic acid sequence which encodes a maize or soybean porphobilinogen synthase enzyme or fragment thereof, a nucleic acid sequence which encodes a maize or soybean hydroxymethylbilane synthase enzyme or fragment thereof, a nucleic acid sequence which encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof, a nucleic acid sequence which encodes a maize or soybean glutamate tRNA ligase enzyme or fragment thereof, a nucleic acid sequence which encodes a maize or soybean glutamyl-tRNA reductase enzyme or fragment thereof, a nucleic acid sequence which encodes a maize or soybean Mg-chelatase enzyme or fragment thereof, and a nucleic acid sequence which encodes a maize or soybean ferrochelatase enzyme or fragment thereof; which is linked to (C) a 3' non-translated sequence that functions in the plant cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule.

The present invention also provides a transformed plant having a nucleic acid molecule which comprises: (A) an exogenous promoter region which functions in a plant cell to cause the production of a mRNA molecule; which is linked to (B) a transcribed nucleic acid molecule with a transcribed strand and a non-transcribed strand, wherein the transcribed strand is complementary to a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or fragment thereof; which is

linked to (C) a 3' non-translated sequence that functions in plant cells to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule.

The present invention also provides a transformed plant having a nucleic acid molecule which comprises: (A) an exogenous promoter region which functions in a plant cell to cause the production of a mRNA molecule; which is linked to: (B) a transcribed nucleic acid molecule with a transcribed strand and a non-transcribed strand, wherein a transcribed mRNA of the transcribed strand is complementary to an endogenous mRNA molecule having a nucleic acid sequence selected from the group consisting of an endogenous mRNA molecule that encodes a putative maize or soybean putative chlorophyll synthetase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean porphobilinogen synthase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean hydromethylbilane synthase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or fragment thereof an endogenous mRNA molecule that encodes a maize

093318-012099

or soybean glutamyl-tRNA reductase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean Mg-chelatase enzyme or fragment thereof and an endogenous mRNA molecule that encodes a maize or soybean ferrochelatase enzyme or fragment thereof; which is linked to (C) a 3' non-translated sequence that functions in the plant cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule.

The present invention also provides a method for determining a level or pattern of a plant tetrapyrrole pathway protein in a plant cell or plant tissue comprising: (A) incubating, under conditions permitting nucleic acid hybridization, a marker nucleic acid molecule, the marker nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof or fragment of either, with a complementary nucleic acid molecule obtained from the plant cell or plant tissue, wherein nucleic acid hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant cell or plant tissue permits the detection of the plant tetrapyrrole pathway protein; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant cell or plant tissue; and (C) detecting the level or pattern of the complementary nucleic acid, wherein the detection of the complementary nucleic acid is predictive of the level or pattern of the plant tetrapyrrole pathway protein.

The present invention also provides a method for determining a level or pattern of a plant tetrapyrrole pathway protein in a plant cell or plant tissue comprising: (A) incubating, under conditions permitting nucleic acid hybridization, a marker nucleic acid molecule, the marker nucleic acid molecule comprising a nucleic acid molecule that encodes a putative maize or

soybean chlorophyll synthetase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or complement thereof or fragment of either and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or complement thereof or fragment of either, with a complementary nucleic acid molecule obtained from the plant cell or plant tissue, wherein nucleic acid hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant cell or plant

tissue permits the detection of the plant tetrapyrrole pathway protein; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant cell or plant tissue; and (C) detecting the level or pattern of the complementary nucleic acid, wherein the detection of the complementary nucleic acid is predictive of the level or pattern of the plant tetrapyrrole pathway protein.

The present invention also provides a method for determining a level or pattern of a plant tetrapyrrole pathway protein in a plant cell or plant tissue under evaluation which comprises assaying the concentration of a molecule, whose concentration is dependent upon the expression of a gene, the gene specifically hybridizes to a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof, in comparison to the concentration of that molecule present in a reference plant cell or a reference plant tissue with a known level or pattern of the plant tetrapyrrole pathway protein, wherein the assayed concentration of the molecule is compared to the assayed concentration of the molecule in the reference plant cell or reference plant tissue with the known level or pattern of the plant tetrapyrrole pathway protein.

The present invention also provides a method for determining a level or pattern of a plant tetrapyrrole pathway protein in a plant cell or plant tissue under evaluation which comprises assaying the concentration of a molecule, whose concentration is dependent upon the expression of a gene, the gene specifically hybridizes to a nucleic acid molecule selected from the group consisting of a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or complement thereof, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or complement

thereof, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or complement thereof, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or complement thereof and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or complement thereof, in comparison to the concentration of that molecule present in a reference plant cell or a reference plant tissue with a known level or pattern of the plant tetrapyrrole pathway protein, wherein the assayed concentration of the molecule is compared to the assayed concentration of the molecule in the reference plant cell or the reference plant tissue with the known level or pattern of the plant tetrapyrrole pathway protein.

The present invention provides a method of determining a mutation in a plant whose presence is predictive of a mutation affecting a level or pattern of a protein comprising the steps: (A) incubating, under conditions permitting nucleic acid hybridization, a marker nucleic acid, the marker nucleic acid selected from the group of marker nucleic acid molecules which specifically

hybridize to a nucleic acid molecule having a nucleic acid sequence selected from the group of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof or fragment of either and a complementary nucleic acid molecule obtained from the plant, wherein nucleic acid hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant permits the detection of a polymorphism whose presence is predictive of a mutation affecting the level or pattern of the protein in the plant; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant; and (C) detecting the presence of the polymorphism, wherein the detection of the polymorphism is predictive of the mutation.

The present invention also provides a method for determining a mutation in a plant whose presence is predictive of a mutation affecting the level or pattern of a plant tetrapyrrole pathway protein comprising the steps: (A) incubating, under conditions permitting nucleic acid hybridization, a marker nucleic acid molecule, the marker nucleic acid molecule comprising a nucleic acid molecule that is linked to a gene, the gene specifically hybridizes to a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof and a complementary nucleic acid molecule obtained from the plant, wherein nucleic acid hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant permits the detection of a polymorphism whose presence is predictive of a mutation affecting the level or pattern of the plant tetrapyrrole pathway protein in the plant; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant; and (C) detecting the presence of the polymorphism, wherein the detection of the polymorphism is predictive of the mutation.



The present invention also provides a method for determining a mutation in a plant whose presence is predictive of a mutation affecting the level or pattern of a plant tetrapyrrole pathway protein comprising the steps: (A) incubating, under conditions permitting nucleic acid hybridization, a marker nucleic acid molecule, the marker nucleic acid molecule comprising a nucleic acid molecule that is linked to a gene, the gene specifically hybridizes to a nucleic acid molecule selected from the group consisting of a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or complement thereof, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or complement thereof, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or complement thereof, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or complement thereof and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or complement

thereof and a complementary nucleic acid molecule obtained from the plant, wherein nucleic acid hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant permits the detection of a polymorphism whose presence is predictive of a mutation affecting the level or pattern of the plant tetrapyrrole pathway protein in the plant; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant; and (C) detecting the presence of the polymorphism, wherein the detection of the polymorphism is predictive of the mutation.

The present invention also provides a method of producing a plant containing an overexpressed protein comprising: (A) transforming the plant with a functional nucleic acid molecule, wherein the functional nucleic acid molecule comprises a promoter region, wherein the promoter region is linked to a structural region, wherein the structural region has a nucleic acid sequence selected from group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 wherein the structural region is linked to a 3' non-translated sequence that functions in the plant to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a mRNA molecule; and wherein the functional nucleic acid molecule results in overexpression of the protein; and (B) growing the transformed plant.

The present invention also provides a method of producing a plant containing an overexpressed plant tetrapyrrole pathway protein comprising: (A) transforming the plant with a functional nucleic acid molecule, wherein the functional nucleic acid molecule comprises a promoter region, wherein the promoter region is linked to a structural region, wherein the structural region comprises a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or fragment thereof; wherein the structural region is linked to a 3' non-translated sequence that functions in the plant to cause

termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a mRNA molecule; and wherein the functional nucleic acid molecule results in overexpression of the plant tetrapyrrole pathway protein; and (B) growing the transformed plant.

The present invention also provides a method of producing a plant containing an overexpressed plant tetrapyrrole pathway protein comprising: (A) transforming the plant with a functional nucleic acid molecule, wherein the functional nucleic acid molecule comprises a promoter region, wherein the promoter region is linked to a structural region, wherein the structural region comprises a nucleic acid molecule selected from the group consisting of a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean

glutamate tRNA ligase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or complement thereof or fragment and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or complement thereof or fragment of either, wherein the structural region is linked to a 3' non-translated sequence that functions in the plant to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a mRNA molecule; and wherein the functional nucleic acid molecule results in overexpression of the plant tetrapyrrole pathway protein; and (B) growing the transformed plant.

The present invention also provides a method of producing a plant containing reduced levels of a plant tetrapyrrole pathway protein comprising: (A) transforming the plant with a functional nucleic acid molecule, wherein the functional nucleic acid molecule comprises a promoter region, wherein the promoter region is linked to a structural region, wherein the structural region comprises a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677; wherein the structural region is linked to a 3' non-translated sequence that functions in the plant to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a mRNA molecule; and wherein the functional nucleic acid molecule results in co-suppression of the plant tetrapyrrole pathway protein; and (B) growing the transformed plant.

The present invention also provides a method of producing a plant containing reduced levels of a plant tetrapyrrole pathway protein comprising: (A) transforming the plant with a functional nucleic acid molecule, wherein the functional nucleic acid molecule comprises a promoter region, wherein the promoter region is linked to a structural region, wherein the

structural region comprises a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or complement thereof or fragment, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or complement thereof or fragment and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or complement thereof or fragment of either, wherein the structural region is linked to a 3' non-translated sequence that functions in the plant to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a

mRNA molecule; and wherein the functional nucleic acid molecule results in co-suppression of the plant tetrapyrrole pathway protein; and (B) growing the transformed plant.

The present invention also provides a method for reducing expression of a plant tetrapyrrole pathway protein in a plant comprising: (A) transforming the plant with a nucleic acid molecule, the nucleic acid molecule having an exogenous promoter region which functions in a plant cell to cause the production of a mRNA molecule, wherein the exogenous promoter region is linked to a transcribed nucleic acid molecule having a transcribed strand and a non-transcribed strand, wherein the transcribed strand is complementary to a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof or fragments of either and the transcribed strand is complementary to an endogenous mRNA molecule; and wherein the transcribed nucleic acid molecule is linked to a 3' non-translated sequence that functions in the plant cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a mRNA molecule; and (B) growing the transformed plant.

The present invention also provides a method for reducing expression of a plant tetrapyrrole pathway protein in a plant comprising: (A) transforming the plant with a nucleic acid molecule, the nucleic acid molecule having an exogenous promoter region which functions in a plant cell to cause the production of a mRNA molecule, wherein the exogenous promoter region is linked to a transcribed nucleic acid molecule having a transcribed strand and a non-transcribed strand, wherein a transcribed mRNA of the transcribed strand is complementary to a nucleic acid molecule selected from the group consisting of an endogenous mRNA molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or

fragment thereof, an endogenous mRNA molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean porphobilinogen synthase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean hydromethylase synthase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or fragment thereof, an endogenous mRNA molecule that encodes a maize or soybean Mg-chelatase enzyme or fragment thereof and an endogenous mRNA molecule that encodes a maize or soybean ferrochelatase enzyme or fragment thereof, and wherein the transcribed nucleic acid molecule is linked to a 3' non-translated sequence that functions in the plant cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of a mRNA molecule; and (B) growing the transformed plant.

The present invention also provides a method of determining an association between a polymorphism and a plant trait comprising: (A) hybridizing a nucleic acid molecule specific for the polymorphism to genetic material of a plant, wherein the nucleic acid molecule has a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or

complements thereof or fragment of either; and (B) calculating the degree of association between the polymorphism and the plant trait.

The present invention also provides a method of determining an association between a polymorphism and a plant trait comprising: (A) hybridizing a nucleic acid molecule specific for the polymorphism to genetic material of a plant, wherein the nucleic acid molecule is selected from the group consisting of a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or complement thereof or fragment of either, nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or



complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or complement thereof or fragment of either and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or complement thereof or fragment of either and (B) calculating the degree of association between the polymorphism and the plant trait.

The present invention also provides a method of isolating a nucleic acid that encodes a plant tetrapyrrole pathway protein or fragment thereof comprising: (A) incubating under conditions permitting nucleic acid hybridization, a first nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof or fragment of either with a complementary second nucleic acid molecule obtained from a plant cell or plant tissue; (B) permitting hybridization between the first nucleic acid molecule and the second nucleic acid molecule obtained from the plant cell or plant tissue; and (C) isolating the second nucleic acid molecule.

The present invention also provides a method of isolating a nucleic acid molecule that encodes a plant tetrapyrrole pathway protein or fragment thereof comprising: (A) incubating under conditions permitting nucleic acid hybridization, a first nucleic acid molecule selected from the group consisting of a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or

soybean protoporphyrinogen oxidase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or complement thereof or fragment of either and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or complement thereof or fragment of either, with a complementary second nucleic acid molecule obtained from a plant cell or plant tissue; (B) permitting hybridization between the plant tetrapyrrole pathway protein nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant cell or plant tissue; and (C) isolating the second nucleic acid molecule.

## **DETAILED DESCRIPTION OF THE INVENTION**

### **Definitions and Agents of the Present Invention**

#### **Definitions:**

As used herein, a tetrapyrrole pathway enzyme is any molecule that is associated with the biosynthesis or degradation of tetrapyrroles.

As used herein, ALA refers to 5-aminolevulinic acid and 4-aminolevulinate.

As used herein, ALA synthase (E.C. 2.3.1.37) refers to any enzyme that catalyzes the condensation of glycine with succinyl-CoA to generate ALA.

As used herein, glutamyl-tRNA synthetase (GluRS) (E.C. 6.1.1.17) refers to any enzyme that converts glutamate to glutamyl-tRNA (Glu-tRNA).

As used herein, glutamyl-tRNA reductase (GluTR) refers to any enzyme that catalyzes the NADPH dependent reduction of Glu-tRNA to glutamate 1-semialdehyde (GSA) with the release of intact tRNA.

As used herein, glutamate-1-semialdehyde aminotransferase (GSA-AT) (E.C. 5.4.3.8) refers to any enzyme that catalyzes the conversion of GSA to ALA

As used herein, porphobilinogen synthase (ALA dehydratase) (E.C. 4.2.1.24) refers to any enzyme that catalyzes the asymmetric condensation of two molecules of ALA to yield porphobilinogen.

As used herein, porphobilinogen deaminase (hydroxymethylbilane synthase) (E.C. 4.3.1.8) refers to any enzyme that catalyzes the formation of the linear tetrapyrrole hydroxymethylbilane.

As used herein, uroporphyrinogen III (co)synthase (E.C. 4.2.1.75) refers to any enzyme that catalyzes the ring closure of the unstable linear tetrapyrrole hydroxymethylbilane and the simultaneous isomerization of the acetyl and propionyl groups at pyrrole ring D forming uroporphyrinogen III.

As used herein, uroporphyrinogen III decarboxylase (E.C. 4.1.1.37) refers to any enzyme that catalyzes the stepwise decarboxylation of all four acetate side chains of uroporphyrinogen III starting with ring D followed by rings A, B, and C respectively to form coproporphyrinogen III.

As used herein, coproporphyrinogen III oxidase (E.C. 1.3.3.3) refers to any enzyme that catalyzes the oxygen dependent sequential oxidative decarboxylation of the A and B propionyl side chains of coproporphyrinogen III to yield two vinyl groups and protoporphyrinogen IX.

As used herein, protoporphyrinogen IX oxidase (E.C. 1.3.3.4) refers to any enzyme that catalyzes the formation of the aromatic protoporphyrin IX by the six electron oxidation of protoporphyrinogen IX.

As used herein, Mg-chelatase refers to any enzyme that catalyzes the conversion of protoporphyrin IX to magnesium protoporphyrin IX by the insertion  $Mg^{+2}$ .

As used herein, Mg-protoporphyrin IX *O*-methyltransferase (E.C. 2.1.1.11) refers to any enzyme that esterifies the propionic side chain of ring III of Mg-protoporphyrin IX to form Mg-protoporphyrin IX monomethylester.

As used herein, Mg-protoporphyrin IX monomethyl ester cyclase refers to any enzyme that catalyzes the cyclization of Mg-protoporphyrin IX monomethylester to form the isocyclic ring E of divinyl protochlorophyllide.

As used herein, 8-vinyl reductase refers to any enzyme that can reduce divinyl protochlorophyllide or Mg-protoporphyrin IX monomethylester to monovinyl protochlorophyllide.

As used herein, protochlorophyllide reductase ("POR") (E.C. 1.3.1.33) refers to any enzyme that catalyzes the reduction of the double bond between carbons 7 and 8 of the D ring of protochlorophyllide producing chlorophyllide

As used herein, chlorophyll synthetase refers to any enzyme that esterifies the propionic acid side chain of ring D of chlorophyllide with either phytyl pyrophosphate or geranylgeranyl pyrophosphate.

As used herein, ferrochelatase (E.C. 4.99.1.1) refers to any enzyme that catalyzes the conversion of protoporphyrin IX to heme.

## **Agents**

### **(a) Nucleic Acid Molecules**

Agents of the present invention include plant nucleic acid molecules and more preferably include maize, soybean and *Arabidopsis thaliana* nucleic acid molecules and more preferably include nucleic acid molecules of the maize genotypes B73 (Illinois Foundation Seeds, Champaign, Illinois U.S.A.), B73 x Mo17 (Illinois Foundation Seeds, Champaign, Illinois U.S.A.), DK604 (Dekalb Genetics, Dekalb, Illinois U.S.A.), H99 (Illinois Foundation Seeds, Champaign, Illinois U.S.A.), RX601 (Asgrow Seed Company, Des Moines, Iowa), Mo17 (Illinois Foundation Seeds, Champaign, Illinois U.S.A.), and soybean types Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa), C1944 (United States Department of Agriculture (USDA) Soybean Germplasm Collection, Urbana, Illinois U.S.A.), Cristalina (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.), FT108 (Monsoy, Brazil), Hartwig (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.), BW211S Null (Tohoku University, Morioka, Japan), PI507354 (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.), Asgrow A4922 (Asgrow Seed Company, Des Moines, Iowa U.S.A.), PI227687 (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.), PI229358 (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) and Asgrow A3237 (Asgrow Seed Company, Des Moines, Iowa U.S.A.).

A subset of the nucleic acid molecules of the present invention includes nucleic acid molecules that are marker molecules. Another subset of the nucleic acid molecules of the present

invention include nucleic acid molecules that encode a protein or fragment thereof. Another subset of the nucleic acid molecules of the present invention are EST molecules.

Fragment nucleic acid molecules may encode significant portion(s) of, or indeed most of, these nucleic acid molecules. Alternatively, the fragments may comprise smaller oligonucleotides (having from about 15 to about 250 nucleotide residues and more preferably, about 15 to about 30 nucleotide residues).

As used herein, an agent, be it a naturally occurring molecule or otherwise may be “substantially purified,” if desired, such that one or more molecules that is or may be present in a naturally occurring preparation containing that molecule will have been removed or will be present at a lower concentration than that at which it would normally be found.

The agents of the present invention will preferably be “biologically active” with respect to either a structural attribute, such as the capacity of a nucleic acid to hybridize to another nucleic acid molecule, or the ability of a protein to be bound by an antibody (or to compete with another molecule for such binding). Alternatively, such an attribute may be catalytic and thus involve the capacity of the agent to mediate a chemical reaction or response.

The agents of the present invention may also be recombinant. As used herein, the term recombinant means any agent (e.g. DNA, peptide etc.), that is, or results, however indirect, from human manipulation of a nucleic acid molecule.

It is understood that the agents of the present invention may be labeled with reagents that facilitate detection of the agent (e.g. fluorescent labels, Prober *et al.*, *Science* 238:336-340 (1987); Albarella *et al.*, EP 144914; chemical labels, Sheldon *et al.*, U.S. Patent 4,582,789; Albarella *et al.*, U.S. Patent 4,563,417; modified bases, Miyoshi *et al.*, EP 119448, all of which are hereby incorporated by reference in their entirety).

It is further understood, that the present invention provides recombinant bacterial, mammalian, microbial, insect, fungal and plant cells and viral constructs comprising the agents of the present invention. (See, for example, Uses of the Agents of the Invention, Section (a) Plant Constructs and Plant Transformants; Section (b) Fungal Constructs and Fungal Transformants; Section (c) Mammalian Constructs and Transformed Mammalian Cells; Section (d) Insect Constructs and Transformed Insect Cells; and Section (e) Bacterial Constructs and Transformed Bacterial Cells)

Nucleic acid molecules or fragments thereof of the present invention are capable of specifically hybridizing to other nucleic acid molecules under certain circumstances. As used herein, two nucleic acid molecules are said to be capable of specifically hybridizing to one another if the two molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure. A nucleic acid molecule is said to be the "complement" of another nucleic acid molecule if they exhibit complete complementarity. As used herein, molecules are said to exhibit "complete complementarity" when every nucleotide of one of the molecules is complementary to a nucleotide of the other. Two molecules are said to be "minimally complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under at least conventional "low-stringency" conditions. Similarly, the molecules are said to be "complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under conventional "high-stringency" conditions. Conventional stringency conditions are described by Sambrook *et al.*, *Molecular Cloning*, A Laboratory Manual, 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York (1989) and by Haymes *et al.*, *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985), the entirety of which is herein incorporated by

reference. Departures from complete complementarity are therefore permissible, as long as such departures do not completely preclude the capacity of the molecules to form a double-stranded structure. Thus, in order for a nucleic acid molecule to serve as a primer or probe it need only be sufficiently complementary in sequence to be able to form a stable double-stranded structure under the particular solvent and salt concentrations employed.

Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 X sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 X SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 X SSC at 50°C to a high stringency of about 0.2 X SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Both temperature and salt may be varied, or either the temperature or the salt concentration may be held constant while the other variable is changed.

In a preferred embodiment, a nucleic acid of the present invention will specifically hybridize to one or more of the nucleic acid molecules set forth in SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof under moderately stringent conditions, for example at about 2.0 X SSC and about 65°C.

In a particularly preferred embodiment, a nucleic acid of the present invention will include those nucleic acid molecules that specifically hybridize to one or more of the nucleic acid molecules set forth in SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof under high stringency conditions such as 0.2 X SSC and about 65°C.



In one aspect of the present invention, the nucleic acid molecules of the present invention have one or more of the nucleic acid sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof. In another aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between 100% and 90% sequence identity with one or more of the nucleic acid sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof. In a further aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between 100% and 95% sequence identity with one or more of the nucleic acid sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof. In a more preferred aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between 100% and 98% sequence identity with one or more of the nucleic acid sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof. In an even more preferred aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between 100% and 99% sequence identity with one or more of the sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof.

In a further more preferred aspect of the present invention, one or more of the nucleic acid molecules of the present invention exhibit 100% sequence identity with a nucleic acid molecule present within MONN01, SATMON001, SATMON003 through SATMON014, SATMON016 through SATMON031, SATMON033, SATMON034, SATMON~001, SATMONN01, SATMONN04 through SATMONN006, CMz029 through CMz031, CMz033 through CMz037, CMz039 through CMz042, CMz044 through CMz045, CMz047 through CMz050, SOYMON001 through SOYMON038, Soy51 through Soy56, Soy58 through Soy62,

Soy65 through Soy73 and Soy76 through Soy77, Lib9, Lib22 through Lib25, Lib35, and Lib146 (Monsanto Company, St. Louis, Missouri U.S.A.).

**(i) Nucleic Acid Molecules Encoding Proteins or Fragments Thereof**

Nucleic acid molecules of the present invention can comprise sequences that encode a tetrapyrrole pathway enzyme or fragment thereof. Such transcription factors or fragments thereof include homologues of known transcription factors in other organisms.

In a preferred embodiment of the present invention, a maize or soybean tetrapyrrole pathway enzyme or fragment thereof of the present invention is a homologue of another plant tetrapyrrole pathway protein. In another preferred embodiment of the present invention, a maize or soybean tetrapyrrole pathway enzyme or fragment thereof of the present invention is a homologue of a fungal tetrapyrrole pathway enzyme. In another preferred embodiment of the present invention, a maize or soybean tetrapyrrole pathway enzyme of the present invention is a homologue of mammalian transcription factor. In another preferred embodiment of the present invention, a maize or soybean tetrapyrrole pathway enzyme or fragment thereof of the present invention is a homologue of a bacterial transcription factor. In another preferred embodiment of the present invention, a maize or soybean tetrapyrrole pathway enzyme or fragment thereof of the present invention is a homologue of a maize tetrapyrrole pathway enzyme. In another preferred embodiment of the present invention, a maize or soybean tetrapyrrole pathway enzyme homologue or fragment thereof of the present invention is a homologue of a soybean transcription factor.

In a preferred embodiment of the present invention, the nucleic molecule of the present invention encodes a maize or soybean tetrapyrrole pathway enzyme or fragment thereof where a maize or soybean tetrapyrrole pathway enzyme exhibits a BLAST probability score of greater

than 1E-12, preferably a BLAST probability score of between about 1E-30 and about 1E-12, even more preferably a BLAST probability score of greater than 1E-30 with its homologue.

In another preferred embodiment of the present invention, the nucleic acid molecule encoding a maize or soybean tetrapyrrole pathway enzyme or fragment thereof exhibits a % identity with its homologue of between about 25% and about 40%, more preferably of between about 40 and about 70%, even more preferably of between about 70% and about 90% and even more preferably between about 90% and 99%. In another preferred embodiment, of the present invention, a maize or soybean tetrapyrrole enzyme or fragment thereof exhibits a % identity with its homologue of 100%.

In a preferred embodiment of the present invention, the nucleic molecule of the present invention encodes a maize or soybean tetrapyrrole pathway enzyme or fragment thereof where a maize or soybean tetrapyrrole pathway enzyme exhibits a BLAST score of greater than 120, preferably a BLAST score of between about 1450 and about 120, even more preferably a BLAST score of greater than 1450 with its homologue.

Nucleic acid molecules of the present invention also include non-maize, non-soybean homologues. Preferred non- homologues are selected from the group consisting of alfalfa, *Arabidopsis*, barley, *Brassica*, broccoli, cabbage, citrus, cotton, garlic, oat, oilseed rape, onion, canola, flax, an ornamental plant, pea, peanut, pepper, potato, rice, rye, sorghum, strawberry, sugarcane, sugarbeet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, oil palm and *Phaseolus*.

In a preferred embodiment, nucleic acid molecules having SEQ ID NO: 1 through SEQ ID NO: 677 or complements and fragments of either can be utilized to obtain such homologues.

The degeneracy of the genetic code, which allows different nucleic acid sequences to code for the same protein or peptide, is known in the literature. (U.S. Patent No. 4,757,006, the entirety of which is herein incorporated by reference).

In an aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding a maize or soybean tetrapyrrole pathway enzyme or fragment thereof in SEQ ID NO: 1 through SEQ ID NO: 677 due to the degeneracy in the genetic code in that they encode the same transcription factor but differ in nucleic acid sequence.

In another further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding a maize or soybean tetrapyrrole pathway enzyme or fragment thereof in SEQ ID NO: 1 through SEQ ID NO: 677 due to fact that the different nucleic acid sequence encodes a transcription factor having one or more conservative amino acid residue. Examples of conservative substitutions are set forth in Table 1. It is understood that codons capable of coding for such conservative substitutions are known in the art.

**Table 1**

<u>Original Residue</u>	<u>Conservative Substitutions</u>
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Cys	Ser; Ala

Gln	Asn
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu; Val
Leu	Ile; Val
Lys	Arg; Gln; Glu
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

In a further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding a maize or soybean tetrapyrrole or fragment thereof set forth in SEQ ID NO: 1 through SEQ ID NO: 677 or fragment thereof due to the fact that one or more codons encoding an amino acid has been substituted for a codon that encodes a nonessential substitution of the amino acid originally encoded.

Agents of the present invention include nucleic acid molecules that encode a maize, or soybean tetrapyrrole pathway enzyme or fragment thereof and particularly substantially purified nucleic acid molecules selected from the group consisting of a nucleic acid molecule that

encodes a putative maize or soybean chlorophyll synthetase enzyme or fragment, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or fragment, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or fragment, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or fragment, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or fragment, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or fragment, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or fragment, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or fragment, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or fragment, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme fragment, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or fragment, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or fragment and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or fragment.

Non-limiting examples of such nucleic acid molecules of the present invention are nucleic acid molecules comprising: SEQ ID NO: 1 through SEQ ID NO: 677 or fragment thereof that encode for a plant tetrapyrrole pathway protein or fragment thereof, SEQ ID NO: 1 through SEQ ID NO: 8 and SEQ ID NO: 384 through SEQ ID NO: 397 or fragment thereof that encode for a putative chlorophyll synthetase enzyme or fragment thereof, SEQ ID NO: 9 through SEQ ID NO: 94 and SEQ ID NO: 398 through SEQ ID NO: 466 or fragment thereof that encode for a protochlorophyllide reductase enzyme or fragment thereof, SEQ ID NO: 95 through SEQ ID NO:

96 and SEQ ID NO: 467 through SEQ ID NO: 479 or fragment thereof that encode for a putative protochlorophyllide reductase enzyme or fragment thereof, SEQ ID NO: 97 through SEQ ID NO: 128 and SEQ ID NO: 480 through SEQ ID NO: 494 or fragment thereof that encodes for a coproporphyrinogen oxidase enzyme or fragment thereof, SEQ ID NO: 129 through SEQ ID NO: 131 and SEQ ID NO: 495 through SEQ ID NO: 499 or fragment thereof that encode for a protoporphyrinogen oxidase enzyme or fragment thereof, SEQ ID NO: 132 through SEQ ID NO: 144 and SEQ ID NO: 500 through SEQ ID NO: 509 or fragment thereof that encode for an uroporphyrinogen decarboxylase enzyme or fragment thereof, SEQ ID NO: 510 or fragment thereof that encode for a putative uroporphyrinogen decarboxylase enzyme or fragment thereof, SEQ ID NO: 145 through SEQ ID NO: 191 and SEQ ID NO: 511 through SEQ ID NO: 531 or fragment thereof that encode for a porphobilinogen synthase enzyme or fragment thereof, SEQ ID NO: 154, SEQ ID NO: 192 through SEQ ID NO: 217 and SEQ ID NO: 532 through SEQ ID NO: 542 or fragment thereof that encode for a hydroxymethylbilane synthase enzyme or fragment thereof, SEQ ID NO: 218 through SEQ ID NO: 265 and SEQ ID NO: 543 through SEQ ID NO: 569 or fragment thereof that encodes for a glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof, SEQ ID NO: 266 through SEQ ID NO: 289 and SEQ ID NO: 570 through SEQ ID NO: 585 or fragment thereof that encode for a glutamate tRNA ligase enzyme or fragment thereof, SEQ ID NO: 290 through SEQ ID NO: 306 and SEQ ID NO: 586 through SEQ ID NO: 609 or fragment thereof that encode for an glutamyl-tRNA reductase enzyme or fragment thereof, SEQ ID NO: 307 through SEQ ID NO: 371 and SEQ ID NO: 610 through SEQ ID NO: 652 or fragment thereof that encode for a Mg-chelatase enzyme or fragment thereof, and SEQ ID NO: 372 through SEQ ID NO: 383 and SEQ ID NO: 653 through SEQ ID NO: 677 or fragment thereof that encode for an ferrochelatase enzyme or fragment thereof.

A nucleic acid molecule of the present invention can also encode an homologue of a putative maize or soybean chlorophyll synthetase enzyme or fragment thereof, a maize or soybean protochlorophyllide reductase enzyme or fragment thereof, a putative maize or soybean protochlorophyllide reductase enzyme or fragment thereof, a maize or soybean coproporphyrinogen oxidase enzyme or fragment thereof, a maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof, a maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof, a putative maize uroporphyrinogen decarboxylase enzyme or fragment thereof, a maize or soybean porphobilinogen synthase enzyme or fragment thereof, a maize or soybean hydroxymethylbilane synthase enzyme or fragment thereof, a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof, a maize or soybean glutamate tRNA ligase enzyme fragment thereof, a maize or soybean glutamyl-tRNA reductase enzyme or fragment thereof, a maize or soybean Mg-chelatase enzyme or fragment thereof and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or fragment thereof. As used herein a homologue protein molecule or fragment thereof is a counterpart protein molecule or fragment thereof in a second species (*e.g.*, maize chlorophyll synthetase is a homologue of *Arabidopsis*' chlorophyll synthetase).

## **(ii) Nucleic Acid Molecule Markers and Probes**

One aspect of the present invention concerns markers that include nucleic acid molecules SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof or fragments of either that can act as markers or other nucleic acid molecules of the present invention that can act as markers. Genetic markers of the present invention include “dominant” or “codominant” markers “Codominant markers” reveal the presence of two or more alleles (two per diploid individual) at a locus. “Dominant markers” reveal the presence of only a single allele per locus. The presence



of the dominant marker phenotype (e.g., a band of DNA) is an indication that one allele is present in either the homozygous or heterozygous condition. The absence of the dominant marker phenotype (e.g. absence of a DNA band) is merely evidence that “some other” undefined allele is present. In the case of populations where individuals are predominantly homozygous and loci are predominately dimorphic, dominant and codominant markers can be equally valuable. As populations become more heterozygous and multi-allelic, codominant markers often become more informative of the genotype than dominant markers. Marker molecules can be, for example, capable of detecting polymorphisms such as single nucleotide polymorphisms (SNPs).

SNPs are single base changes in genomic DNA sequence. They occur at greater frequency and are spaced with a greater uniformity throughout a genome than other reported forms of polymorphism. The greater frequency and uniformity of SNPs means that there is greater probability that such a polymorphism will be found near or in a genetic locus of interest than would be the case for other polymorphisms. SNPs are located in protein-coding regions and noncoding regions of a genome. Some of these SNPs may result in defective or variant protein expression (e.g., as a results of mutations or defective splicing). Analysis (genotyping) of characterized SNPs can require only a plus/minus assay rather than a lengthy measurement, permitting easier automation.

SNPs can be characterized using any of a variety of methods. Such methods include the direct or indirect sequencing of the site, the use of restriction enzymes (Botstein *et al.*, *Am. J. Hum. Genet.* 32:314-331 (1980), the entirety of which is herein incorporated reference; Konieczny and Ausubel, *Plant J.* 4:403-410 (1993), the entirety of which is herein incorporated by reference), enzymatic and chemical mismatch assays (Myers *et al.*, *Nature* 313:495-498

(1985), the entirety of which is herein incorporated by reference), allele-specific PCR (Newton *et al.*, *Nucl. Acids Res.* 17:2503-2516 (1989), the entirety of which is herein incorporated by reference; Wu *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86:2757-2760 (1989), the entirety of which is herein incorporated by reference), ligase chain reaction (Barany, *Proc. Natl. Acad. Sci. (U.S.A.)* 88:189-193 (1991), the entirety of which is herein incorporated by reference), single-strand conformation polymorphism analysis (Labrune *et al.*, *Am. J. Hum. Genet.* 48: 1115-1120 (1991), the entirety of which is herein incorporated by reference), primer-directed nucleotide incorporation assays (Kuppuswami *et al.*, *Proc. Natl. Acad. Sci. USA* 88:1143-1147 (1991), the entirety of which is herein incorporated by reference), dideoxy fingerprinting (Sarkar *et al.*, *Genomics* 13:441-443 (1992), the entirety of which is herein incorporated by reference), solid-phase ELISA-based oligonucleotide ligation assays (Nikiforov *et al.*, *Nucl. Acids Res.* 22:4167-4175 (1994), the entirety of which is herein incorporated by reference), oligonucleotide fluorescence-quenching assays (Livak *et al.*, *PCR Methods Appl.* 4:357-362 (1995), the entirety of which is herein incorporated by reference), 5'-nuclease allele-specific hybridization TaqMan assay (Livak *et al.*, *Nature Genet.* 9:341-342 (1995), the entirety of which is herein incorporated by reference), template-directed dye-terminator incorporation (TDI) assay (Chen and Kwok, *Nucl. Acids Res.* 25:347-353 (1997), the entirety of which is herein incorporated by reference), allele-specific molecular beacon assay (Tyagi *et al.*, *Nature Biotech.* 16: 49-53 (1998), the entirety of which is herein incorporated by reference), PinPoint assay (Haff and Smirnov, *Genome Res.* 7: 378-388 (1997), the entirety of which is herein incorporated by reference) and dCAPS analysis (Neff *et al.*, *Plant J.* 14:387-392 (1998), the entirety of which is herein incorporated by reference).

Additional markers, such as AFLP markers, RFLP markers and RAPD markers, can be utilized (Walton, *Seed World* 22-29 (July, 1993), the entirety of which is herein incorporated by reference; Burow and Blake, *Molecular Dissection of Complex Traits*, 13-29, Paterson (ed.), CRC Press, New York (1988), the entirety of which is herein incorporated by reference). DNA markers can be developed from nucleic acid molecules using restriction endonucleases, the PCR and/or DNA sequence information. RFLP markers result from single base changes or insertions/deletions. These codominant markers are highly abundant in plant genomes, have a medium level of polymorphism and are developed by a combination of restriction endonuclease digestion and Southern blotting hybridization. CAPS are similarly developed from restriction nuclease digestion but only of specific PCR products. These markers are also codominant, have a medium level of polymorphism and are highly abundant in the genome. The CAPS result from single base changes and insertions/deletions.

Another marker type, RAPDs, are developed from DNA amplification with random primers and result from single base changes and insertions/deletions in plant genomes. They are dominant markers with a medium level of polymorphisms and are highly abundant. AFLP markers require using the PCR on a subset of restriction fragments from extended adapter primers. These markers are both dominant and codominant are highly abundant in genomes and exhibit a medium level of polymorphism.

SSRs require DNA sequence information. These codominant markers result from repeat length changes, are highly polymorphic and do not exhibit as high a degree of abundance in the genome as CAPS, AFLPs and RAPDs SNPs also require DNA sequence information. These codominant markers result from single base substitutions. They are highly abundant and exhibit a medium of polymorphism (Rafalski *et al.*, In: *Nonmammalian Genomic Analysis*, Birren and

Lai (ed.), Academic Press, San Diego, CA, pp. 75-134 (1996), the entirety of which is herein incorporated by reference). It is understood that a nucleic acid molecule of the present invention may be used as a marker.

A PCR probe is a nucleic acid molecule capable of initiating a polymerase activity while in a double-stranded structure to with another nucleic acid. Various methods for determining the structure of PCR probes and PCR techniques exist in the art. Computer generated searches using programs such as Primer3 ([www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi](http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi)), STSPipeline ([www-genome.wi.mit.edu/cgi-bin/www-STS\\_Pipeline](http://www-genome.wi.mit.edu/cgi-bin/www-STS_Pipeline)), or GeneUp (Pesole *et al.*, *BioTechniques* 25:112-123 (1998) the entirety of which is herein incorporated by reference), for example, can be used to identify potential PCR primers.

It is understood that a fragment of one or more of the nucleic acid molecules of the present invention may be a probe and specifically a PCR probe.

#### **(b) Protein and Peptide Molecules**

A class of agents comprises one or more of the protein or fragments thereof or peptide molecules encoded by SEQ ID NO: 1 through SEQ ID NO: 677 or one or more of the protein or fragment thereof and peptide molecules encoded by other nucleic acid agents of the present invention. As used herein, the term "protein molecule" or "peptide molecule" includes any molecule that comprises five or more amino acids. It is well known in the art that proteins may undergo modification, including post-translational modifications, such as, but not limited to, disulfide bond formation, glycosylation, phosphorylation, or oligomerization. Thus, as used herein, the term "protein molecule" or "peptide molecule" includes any protein molecule that is modified by any biological or non-biological process. The terms "amino acid" and "amino

acids" refer to all naturally occurring L-amino acids. This definition is meant to include norleucine, ornithine, homocysteine and homoserine.

Non-limiting examples of the protein or fragment thereof of the present invention include a maize or soybean tetrapyrrole pathway enzyme or fragment thereof, a putative maize or soybean chlorophyll synthetase enzyme or fragment thereof, a maize or soybean protochlorophyllide reductase enzyme or fragment or fragment thereof, a putative maize or soybean protochlorophyllide reductase enzyme or fragment or fragment thereof, a maize or soybean coproporphyrinogen oxidase enzyme or fragment thereof, a maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof, a maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof, a putative maize uroporphyrinogen decarboxylase enzyme or fragment thereof, a maize or soybean porphobilinogen synthase enzyme or fragment thereof, a maize or soybean hydroxymethylbilane synthase enzyme or fragment thereof, a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof, a maize or soybean glutamate tRNA ligase enzyme fragment thereof, a maize or soybean glutamyl-tRNA reductase enzyme or fragment thereof, a maize or soybean Mg-chelatase enzyme or fragment thereof and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or fragment thereof.

Non-limiting examples of the protein or fragment molecules of the present invention are a transcription factor or fragment thereof encoded by: SEQ ID NO: 1 through SEQ ID NO: 677 or fragment thereof that encode for a tetrapyrrole pathway enzyme or fragment thereof, SEQ ID NO: 1 through SEQ ID NO: 8 and SEQ ID NO: 384 through SEQ ID NO: 397 or fragment thereof that encode for a putative chlorophyll synthetase enzyme or fragment thereof, SEQ ID NO: 9 through SEQ ID NO: 94 and SEQ ID NO: 398 through SEQ ID NO: 466 or fragment

thereof that encode for a protochlorophyllide reductase enzyme or fragment thereof, SEQ ID NO: 95 through SEQ ID NO: 96 and SEQ ID NO: 467 through SEQ ID NO: 479 or fragment thereof that encode for a putative protochlorophyllide reductase enzyme or fragment thereof, SEQ ID NO: 97 through SEQ ID NO: 128 and SEQ ID NO: 480 through SEQ ID NO: 494 or fragment thereof that encodes for a coproporphyrinogen oxidase enzyme or fragment thereof, SEQ ID NO: 129 through SEQ ID NO: 131 and SEQ ID NO: 495 through SEQ ID NO: 499 or fragment thereof that encode for a protoporphyrinogen oxidase enzyme or fragment thereof, SEQ ID NO: 132 through SEQ ID NO: 144 and SEQ ID NO: 500 through SEQ ID NO: 509 or fragment thereof that encode for an uroporphyrinogen decarboxylase enzyme or fragment thereof, SEQ ID NO: 510 or fragment thereof that encode for a putative uroporphyrinogen decarboxylase enzyme or fragment thereof, SEQ ID NO: 145 through SEQ ID NO: 191 and SEQ ID NO: 511 through SEQ ID NO: 531 or fragment thereof that encode for a porphobilinogen synthase enzyme or fragment thereof, SEQ ID NO: 154, SEQ ID NO: 192 through SEQ ID NO: 217 and SEQ ID NO: 532 through SEQ ID NO: 542 or fragment thereof that encode for a hydroxymethylbilane synthase enzyme or fragment thereof, SEQ ID NO: 218 through SEQ ID NO: 265 and SEQ ID NO: 543 through SEQ ID NO: 569 or fragment thereof that encodes for a glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof, SEQ ID NO: 266 through SEQ ID NO: 289 and SEQ ID NO: 570 through SEQ ID NO: 585 or fragment thereof that encode for a glutamate tRNA ligase enzyme or fragment thereof, SEQ ID NO: 290 through SEQ ID NO: 306 and SEQ ID NO: 586 through SEQ ID NO: 609 or fragment thereof that encode for an glutamyl-tRNA reductase enzyme or fragment thereof, SEQ ID NO: 307 through SEQ ID NO: 371 and SEQ ID NO: 610 through SEQ ID NO: 652 or fragment thereof that encode for a Mg-chelatase enzyme or fragment thereof, and SEQ ID NO: 372 through SEQ ID NO: 383 and SEQ ID NO:

653 through SEQ ID NO: 677 or fragment thereof that encode for an ferrochelatase enzyme or fragment thereof.

One or more of the protein or fragment of peptide molecules may be produced via chemical synthesis, or more preferably, by expressing in a suitable bacterial or eucaryotic host. Suitable methods for expression are described by Sambrook *et al.*, (In: *Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, New York* (1989)), or similar texts. For example, the protein may be expressed in, for example, Uses of the Agents of the Invention, Section (a) Plant Constructs and Plant Transformants; Section (b) Fungal Constructs and Fungal Transformants; Section (c) Mammalian Constructs and Transformed Mammalian Cells; Section (d) Insect Constructs and Transformed Insect Cells; and Section (e) Bacterial Constructs and Transformed Bacterial Cells.

A “protein fragment” is a peptide or polypeptide molecule whose amino acid sequence comprises a subset of the amino acid sequence of that protein. A protein or fragment thereof that comprises one or more additional peptide regions not derived from that protein is a “fusion” protein. Such molecules may be derivatized to contain carbohydrate or other moieties (such as keyhole limpet hemocyanin, etc.). Fusion protein or peptide molecules of the present invention are preferably produced via recombinant means.

Another class of agents comprise protein or peptide molecules or fragments or fusions thereof encoded by SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof in which conservative, non-essential or non-relevant amino acid residues have been added, replaced or deleted. Computerized means for designing modifications in protein structure are known in the art (Dahiyat and Mayo, *Science* 278:82-87 (1997), the entirety of which is herein incorporated by reference).

The protein molecules of the present invention include plant homologue proteins. An example of such a homologue is a homologue protein of a non-maize or non soybean plant species, that include but not limited to alfalfa, *Arabidopsis*, barley, *Brassica*, broccoli, cabbage, citrus, cotton, garlic, oat, oilseed rape, onion, canola, flax, an ornamental plant, pea, peanut, pepper, potato, rice, rye, sorghum, strawberry, sugarcane, sugarbeet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, oil palm, *Phaseolus* etc. Particularly preferred non-maize or non-soybean for use for the isolation of homologs would include, *Arabidopsis*, barley, cotton, oat, oilseed rape, rice, canola, ornamentals, sugarcane, sugarbeet, tomato, potato, wheat and turf grasses. Such a homologue can be obtained by any of a variety of methods. Most preferably, as indicated above, one or more of the disclosed sequences (SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof) will be used to define a pair of primers that may be used to isolate the homologue-encoding nucleic acid molecules from any desired species. Such molecules can be expressed to yield homologues by recombinant means.

### **(c) Antibodies**

One aspect of the present invention concerns antibodies, single-chain antigen binding molecules, or other proteins that specifically bind to one or more of the protein or peptide molecules of the present invention and their homologues, fusions or fragments. Such antibodies may be used to quantitatively or qualitatively detect the protein or peptide molecules of the present invention. As used herein, an antibody or peptide is said to “specifically bind” to a protein or peptide molecule of the present invention if such binding is not competitively inhibited by the presence of non-related molecules.



Nucleic acid molecules that encode all or part of the protein of the present invention can be expressed, via recombinant means, to yield protein or peptides that can in turn be used to elicit antibodies that are capable of binding the expressed protein or peptide. Such antibodies may be used in immunoassays for that protein. Such protein-encoding molecules, or their fragments may be a "fusion" molecule (i.e., a part of a larger nucleic acid molecule) such that, upon expression, a fusion protein is produced. It is understood that any of the nucleic acid molecules of the present invention may be expressed, via recombinant means, to yield proteins or peptides encoded by these nucleic acid molecules.

The antibodies that specifically bind proteins and protein fragments of the present invention may be polyclonal or monoclonal and may comprise intact immunoglobulins, or antigen binding portions of immunoglobulins fragments (such as  $(F(ab'), F(ab')_2)$ ), or single-chain immunoglobulins producible, for example, via recombinant means. It is understood that practitioners are familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of antibodies (*see*, for example, Harlow and Lane, In: *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1988), the entirety of which is herein incorporated by reference).

Murine monoclonal antibodies are particularly preferred. BALB/c mice are preferred for this purpose, however, equivalent strains may also be used. The animals are preferably immunized with approximately 25  $\mu$ g of purified protein (or fragment thereof) that has been emulsified in a suitable adjuvant (such as TiterMax adjuvant (Vaxcel, Norcross, GA)). Immunization is preferably conducted at two intramuscular sites, one intraperitoneal site and one subcutaneous site at the base of the tail. An additional i.v. injection of approximately 25  $\mu$ g of

antigen is preferably given in normal saline three weeks later. After approximately 11 days following the second injection, the mice may be bled and the blood screened for the presence of anti-protein or peptide antibodies. Preferably, a direct binding Enzyme-Linked Immunoassay (ELISA) is employed for this purpose.

More preferably, the mouse having the highest antibody titer is given a third i.v. injection of approximately 25  $\mu\text{g}$  of the same protein or fragment. The splenic leukocytes from this animal may be recovered 3 days later and then permitted to fuse, most preferably, using polyethylene glycol, with cells of a suitable myeloma cell line (such as, for example, the P3X63Ag8.653 myeloma cell line). Hybridoma cells are selected by culturing the cells under "HAT" (hypoxanthine-aminopterin-thymine) selection for about one week. The resulting clones may then be screened for their capacity to produce monoclonal antibodies ("mAbs"), preferably by direct ELISA.

In one embodiment, anti-protein or peptide monoclonal antibodies are isolated using a fusion of a protein or peptide of the present invention, or conjugate of a protein or peptide of the present invention, as immunogens. Thus, for example, a group of mice can be immunized using a fusion protein emulsified in Freund's complete adjuvant (*e.g.* approximately 50  $\mu\text{g}$  of antigen per immunization). At three week intervals, an identical amount of antigen is emulsified in Freund's incomplete adjuvant and used to immunize the animals. Ten days following the third immunization, serum samples are taken and evaluated for the presence of antibody. If antibody titers are too low, a fourth booster can be employed. Polysera capable of binding the protein or peptide can also be obtained using this method.

In a preferred procedure for obtaining monoclonal antibodies, the spleens of the above-described immunized mice are removed, disrupted and immune splenocytes are isolated over a

ficoll gradient. The isolated splenocytes are fused, using polyethylene glycol with BALB/c-derived HGPRT (hypoxanthine guanine phosphoribosyl transferase) deficient P3x63xAg8.653 plasmacytoma cells. The fused cells are plated into 96 well microtiter plates and screened for hybridoma fusion cells by their capacity to grow in culture medium supplemented with hypoxanthine, aminopterin and thymidine for approximately 2-3 weeks.

Hybridoma cells that arise from such incubation are preferably screened for their capacity to produce an immunoglobulin that binds to a protein of interest. An indirect ELISA may be used for this purpose. In brief, the supernatants of hybridomas are incubated in microtiter wells that contain immobilized protein. After washing, the titer of bound immunoglobulin can be determined using, for example, a goat anti-mouse antibody conjugated to horseradish peroxidase. After additional washing, the amount of immobilized enzyme is determined (for example through the use of a chromogenic substrate). Such screening is performed as quickly as possible after the identification of the hybridoma in order to ensure that a desired clone is not overgrown by non-secreting neighbor cells. Desirably, the fusion plates are screened several times since the rates of hybridoma growth vary. In a preferred embodiment, a different antigenic form may be used to screen the hybridoma. Thus, for example, the splenocytes may be immunized with one immunogen, but the resulting hybridomas can be screened using a different immunogen. It is understood that any of the protein or peptide molecules of the present invention may be used to raise antibodies.

As discussed below, such antibody molecules or their fragments may be used for diagnostic purposes. Where the antibodies are intended for diagnostic purposes, it may be desirable to derivatize them, for example with a ligand group (such as biotin) or a detectable marker group (such as a fluorescent group, a radioisotope or an enzyme).

The ability to produce antibodies that bind the protein or peptide molecules of the present invention permits the identification of mimetic compounds of those molecules. A "mimetic compound" is a compound that is not that compound, or a fragment of that compound, but which nonetheless exhibits an ability to specifically bind to antibodies directed against that compound.

It is understood that any of the agents of the present invention can be substantially purified and/or be biologically active and/or recombinant.

### **Uses of the Agents of the Invention**

Nucleic acid molecules and fragments thereof of the present invention may be employed to obtain other nucleic acid molecules from the same species (e.g., ESTs or fragment thereof from maize may be utilized to obtain other nucleic acid molecules from maize). Such nucleic acid molecules include the nucleic acid molecules that encode the complete coding sequence of a protein and promoters and flanking sequences of such molecules. In addition, such nucleic acid molecules include nucleic acid molecules that encode for other isozymes or gene family members. Such molecules can be readily obtained by using the above-described nucleic acid molecules or fragments thereof to screen cDNA or genomic libraries obtained from maize or soybean. Methods for forming such libraries are well known in the art.

Nucleic acid molecules and fragments thereof of the present invention may also be employed to obtain nucleic acid homologues. Such homologues include the nucleic acid molecule of other plants or other organisms (*e.g.*, alfalfa, *Arabidopsis*, barley, *Brassica*, broccoli, cabbage, citrus, cotton, garlic, oat, oilseed rape, onion, canola, flax, an ornamental plant, pea, peanut, pepper, potato, rice, rye, sorghum, strawberry, sugarcane, sugarbeet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, oil palm, *Phaseolus*, etc.) including the nucleic acid molecules that encode, in whole or in part,

protein homologues of other plant species or other organisms, sequences of genetic elements such as promoters and transcriptional regulatory elements. Such molecules can be readily obtained by using the above-described nucleic acid molecules or fragments thereof to screen cDNA or genomic libraries obtained from such plant species. Methods for forming such libraries are well known in the art. Such homologue molecules may differ in their nucleotide sequences from those found in one or more of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof because complete complementarity is not needed for stable hybridization. The nucleic acid molecules of the present invention therefore also include molecules that, although capable of specifically hybridizing with the nucleic acid molecules may lack "complete complementarity."

Any of a variety of methods may be used to obtain one or more of the above-described nucleic acid molecules (Zamechik *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 83:4143-4146 (1986), the entirety of which is herein incorporated by reference; Goodchild *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 85:5507-5511 (1988), the entirety of which is herein incorporated by reference; Wickstrom *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 85:1028-1032 (1988), the entirety of which is herein incorporated by reference; Holt *et al.*, *Molec. Cell. Biol.* 8:963-973 (1988), the entirety of which is herein incorporated by reference; Gerwirth *et al.*, *Science* 242:1303-1306 (1988), the entirety of which is herein incorporated by reference; Anfossi *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86:3379-3383 (1989), the entirety of which is herein incorporated by reference; Becker *et al.*, *EMBO J.* 8:3685-3691 (1989); the entirety of which is herein incorporated by reference). Automated nucleic acid synthesizers may be employed for this purpose. In lieu of such synthesis, the disclosed nucleic acid molecules may be used to define a pair of primers that can be used with the polymerase chain reaction (Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 51:263-273 (1986); Erlich *et al.*, European Patent 50,424; European Patent 84,796;

European Patent 258,017; European Patent 237,362; Mullis, European Patent 201,184; Mullis *et al.*, U.S. Patent 4,683,202; Erlich, U.S. Patent 4,582,788; and Saiki *et al.*, U.S. Patent 4,683,194, all of which are herein incorporated by reference in their entirety) to amplify and obtain any desired nucleic acid molecule or fragment.

Promoter sequence(s) and other genetic elements, including but not limited to transcriptional regulatory flanking sequences, associated with one or more of the disclosed nucleic acid sequences can also be obtained using the disclosed nucleic acid sequence provided herein. In one embodiment, such sequences are obtained by incubating EST nucleic acid molecules or preferably fragments thereof with members of genomic libraries (*e.g.* maize and soybean) and recovering clones that hybridize to the EST nucleic acid molecule or fragment thereof. In a second embodiment, methods of "chromosome walking," or inverse PCR may be used to obtain such sequences (Frohman *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 85:8998-9002 (1988); Ohara *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86:5673-5677 (1989); Pang *et al.*, *Biotechniques* 22:1046-1048 (1977); Huang *et al.*, *Methods Mol. Biol.* 69:89-96 (1997); Huang *et al.*, *Method Mol. Biol.* 67:287-294 (1997); Benkel *et al.*, *Genet. Anal.* 13:123-127 (1996); Hartl *et al.*, *Methods Mol. Biol.* 58:293-301 (1996), all of which are herein incorporated by reference in their entirety).

The nucleic acid molecules of the present invention may be used to isolate promoters of cell enhanced, cell specific, tissue enhanced, tissue specific, developmentally or environmentally regulated expression profiles. Isolation and functional analysis of the 5' flanking promoter sequences of these genes from genomic libraries, for example, using genomic screening methods and PCR techniques would result in the isolation of useful promoters and transcriptional regulatory elements. These methods are known to those of skill in the art and have been

described (See, for example, Birren *et al.*, *Genome Analysis: Analyzing DNA*, 1, (1997), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., the entirety of which is herein incorporated by reference). Promoters obtained utilizing the nucleic acid molecules of the present invention could also be modified to affect their control characteristics. Examples of such modifications would include but are not limited to enhanced sequences as reported in Uses of the Agents of the Invention, Section (a) Plant Constructs and Plant Transformants. Such genetic elements could be used to enhance gene expression of new and existing traits for crop improvements.

In one sub-aspect, such an analysis is conducted by determining the presence and/or identity of polymorphism(s) by one or more of the nucleic acid molecules of the present invention and more preferably one or more of the EST nucleic acid molecule or fragment thereof which are associated with a phenotype, or a predisposition to that phenotype.

Any of a variety of molecules can be used to identify such polymorphism(s). In one embodiment, one or more of the EST nucleic acid molecules (or a sub-fragment thereof) may be employed as a marker nucleic acid molecule to identify such polymorphism(s). Alternatively, such polymorphisms can be detected through the use of a marker nucleic acid molecule or a marker protein that is genetically linked to (i.e., a polynucleotide that co-segregates with) such polymorphism(s).

In an alternative embodiment, such polymorphisms can be detected through the use of a marker nucleic acid molecule that is physically linked to such polymorphism(s). For this purpose, marker nucleic acid molecules comprising a nucleotide sequence of a polynucleotide located within 1mb of the polymorphism(s) and more preferably within 100kb of the polymorphism(s) and most preferably within 10kb of the polymorphism(s) can be employed.

The genomes of animals and plants naturally undergo spontaneous mutation in the course of their continuing evolution (Gusella, *Ann. Rev. Biochem.* 55:831-854 (1986)). A "polymorphism" is a variation or difference in the sequence of the gene or its flanking regions that arises in some of the members of a species. The variant sequence and the "original" sequence co-exist in the species' population. In some instances, such co-existence is in stable or quasi-stable equilibrium.

A polymorphism is thus said to be "allelic," in that, due to the existence of the polymorphism, some members of a species may have the original sequence (i.e., the original "allele") whereas other members may have the variant sequence (i.e., the variant "allele"). In the simplest case, only one variant sequence may exist and the polymorphism is thus said to be di-allelic. In other cases, the species' population may contain multiple alleles and the polymorphism is termed tri-allelic, etc. A single gene may have multiple different unrelated polymorphisms. For example, it may have a di-allelic polymorphism at one site and a multi-allelic polymorphism at another site.

The variation that defines the polymorphism may range from a single nucleotide variation to the insertion or deletion of extended regions within a gene. In some cases, the DNA sequence variations are in regions of the genome that are characterized by short tandem repeats (STRs) that include tandem di- or tri-nucleotide repeated motifs of nucleotides. Polymorphisms characterized by such tandem repeats are referred to as "variable number tandem repeat" ("VNTR") polymorphisms. VNTRs have been used in identity analysis (Weber, U.S. Patent 5,075,217; Armour *et al.*, *FEBS Lett.* 307:113-115 (1992); Jones *et al.*, *Eur. J. Haematol.* 39:144-147 (1987); Horn *et al.*, PCT Patent Application WO91/14003; Jeffreys, European Patent Application 370,719; Jeffreys, U.S. Patent 5,175,082; Jeffreys *et al.*, *Amer. J. Hum. Genet.*



39:11-24 (1986); Jeffreys *et al.*, *Nature* 316:76-79 (1985); Gray *et al.*, *Proc. R. Acad. Soc. Lond.* 243:241-253 (1991); Moore *et al.*, *Genomics* 10:654-660 (1991); Jeffreys *et al.*, *Anim. Genet.* 18:1-15 (1987); Hillel *et al.*, *Anim. Genet.* 20:145-155 (1989); Hillel *et al.*, *Genet.* 124:783-789 (1990), all of which are herein incorporated by reference in their entirety).

The detection of polymorphic sites in a sample of DNA may be facilitated through the use of nucleic acid amplification methods. Such methods specifically increase the concentration of polynucleotides that span the polymorphic site, or include that site and sequences located either distal or proximal to it. Such amplified molecules can be readily detected by gel electrophoresis or other means.

The most preferred method of achieving such amplification employs the polymerase chain reaction ("PCR") (Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 51:263-273 (1986); Erlich *et al.*, European Patent Appln. 50,424; European Patent Appln. 84,796; European Patent Application 258,017; European Patent Appln. 237,362; Mullis, European Patent Appln. 201,184; Mullis *et al.*, U.S. Patent No. 4,683,202; Erlich, U.S. Patent No. 4,582,788; and Saiki *et al.*, U.S. Patent No. 4,683,194), using primer pairs that are capable of hybridizing to the proximal sequences that define a polymorphism in its double-stranded form.

In lieu of PCR, alternative methods, such as the "Ligase Chain Reaction" ("LCR") may be used (Barany, *Proc. Natl. Acad. Sci. (U.S.A.)* 88:189-193 (1991), the entirety of which is herein incorporated by reference). LCR uses two pairs of oligonucleotide probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependent ligase. As with PCR, the resulting

products thus serve as a template in subsequent cycles and an exponential amplification of the desired sequence is obtained.

LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a polymorphic site. In one embodiment, either oligonucleotide will be designed to include the actual polymorphic site of the polymorphism. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the polymorphic site present on the oligonucleotide. Alternatively, the oligonucleotides may be selected such that they do not include the polymorphic site (see, Segev, PCT Application WO 90/01069, the entirety of which is herein incorporated by reference).

The "Oligonucleotide Ligation Assay" ("OLA") may alternatively be employed (Landegren *et al.*, *Science* 241:1077-1080 (1988), the entirety of which is herein incorporated by reference). The OLA protocol uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target. OLA, like LCR, is particularly suited for the detection of point mutations. Unlike LCR, however, OLA results in "linear" rather than exponential amplification of the target sequence.

Nickerson *et al.*, have described a nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 87:8923-8927 (1990), the entirety of which is herein incorporated by reference). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA. In addition to requiring multiple and separate, processing steps, one problem associated with such combinations is that they inherit all of the problems associated with PCR and OLA.

Schemes based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, are also known (Wu *et al.*, *Genomics* 4:560-569 (1989), the entirety of which is herein incorporated by reference) and may be readily adapted to the purposes of the present invention.

Other known nucleic acid amplification procedures, such as allele-specific oligomers, branched DNA technology, transcription-based amplification systems, or isothermal amplification methods may also be used to amplify and analyze such polymorphisms (Malek *et al.*, U.S. Patent 5,130,238; Davey *et al.*, European Patent Application 329,822; Schuster *et al.*, U.S. Patent 5,169,766; Miller *et al.*, PCT Patent Application WO 89/06700; Kwoh *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86:1173-1177 (1989); Gingeras *et al.*, PCT Patent Application WO 88/10315; Walker *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 89:392-396 (1992), all of which are herein incorporated by reference in their entirety).

The identification of a polymorphism can be determined in a variety of ways. By correlating the presence or absence of it in a plant with the presence or absence of a phenotype, it is possible to predict the phenotype of that plant. If a polymorphism creates or destroys a restriction endonuclease cleavage site, or if it results in the loss or insertion of DNA (e.g., a VNTR polymorphism), it will alter the size or profile of the DNA fragments that are generated by digestion with that restriction endonuclease. As such, individuals that possess a variant sequence can be distinguished from those having the original sequence by restriction fragment analysis. Polymorphisms that can be identified in this manner are termed "restriction fragment length polymorphisms" ("RFLPs"). RFLPs have been widely used in human and plant genetic analyses (Glassberg, UK Patent Application 2135774; Skolnick *et al.*, *Cytogen. Cell Genet.*

32:58-67 (1982); Botstein *et al.*, *Ann. J. Hum. Genet.* 32:314-331 (1980); Fischer *et al.*, (PCT Application WO90/13668); Uhlen, PCT Application WO90/11369).

Polymorphisms can also be identified by Single Strand Conformation Polymorphism (SSCP) analysis. SSCP is a method capable of identifying most sequence variations in a single strand of DNA, typically between 150 and 250 nucleotides in length (Elles, *Methods in Molecular Medicine: Molecular Diagnosis of Genetic Diseases*, Humana Press (1996), the entirety of which is herein incorporated by reference); Orita *et al.*, *Genomics* 5:874-879 (1989), the entirety of which is herein incorporated by reference). Under denaturing conditions a single strand of DNA will adopt a conformation that is uniquely dependent on its sequence conformation. This conformation usually will be different, even if only a single base is changed. Most conformations have been reported to alter the physical configuration or size sufficiently to be detectable by electrophoresis. A number of protocols have been described for SSCP including, but not limited to, Lee *et al.*, *Anal. Biochem.* 205:289-293 (1992), the entirety of which is herein incorporated by reference; Suzuki *et al.*, *Anal. Biochem.* 192:82-84 (1991), the entirety of which is herein incorporated by reference; Lo *et al.*, *Nucleic Acids Research* 20:1005-1009 (1992), the entirety of which is herein incorporated by reference; Sarkar *et al.*, *Genomics* 13:441-443 (1992), the entirety of which is herein incorporated by reference. It is understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by SSCP analysis.

Polymorphisms may also be found using a DNA fingerprinting technique called amplified fragment length polymorphism (AFLP), which is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA to profile that DNA (Vos *et al.*, *Nucleic Acids Res.* 23:4407-4414 (1995), the entirety of which is herein incorporated

by reference). This method allows for the specific co-amplification of high numbers of restriction fragments, which can be visualized by PCR without knowledge of the nucleic acid sequence.

AFLP employs basically three steps. Initially, a sample of genomic DNA is cut with restriction enzymes and oligonucleotide adapters are ligated to the restriction fragments of the DNA. The restriction fragments are then amplified using PCR by using the adapter and restriction sequence as target sites for primer annealing. The selective amplification is achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotide flanking the restriction sites. These amplified fragments are then visualized on a denaturing polyacrylamide gel.

AFLP analysis has been performed on *Salix* (Beismann *et al.*, *Mol. Ecol.* 6:989-993 (1997), the entirety of which is herein incorporated by reference), *Acinetobacter* (Janssen *et al.*, *Int. J. Syst. Bacteriol.* 47:1179-1187 (1997), the entirety of which is herein incorporated by reference), *Aeromonas popoffi* (Huys *et al.*, *Int. J. Syst. Bacteriol.* 47:1165-1171 (1997), the entirety of which is herein incorporated by reference), rice (McCouch *et al.*, *Plant Mol. Biol.* 35:89-99 (1997), the entirety of which is herein incorporated by reference; Nandi *et al.*, *Mol. Gen. Genet.* 255:1-8 (1997), the entirety of which is herein incorporated by reference; Cho *et al.*, *Genome* 39:373-378 (1996), the entirety of which is herein incorporated by reference), barley (*Hordeum vulgare*)(Simons *et al.*, *Genomics* 44:61-70 (1997), the entirety of which is herein incorporated by reference; Waugh *et al.*, *Mol. Gen. Genet.* 255:311-321 (1997), the entirety of which is herein incorporated by reference; Qi *et al.*, *Mol. Gen. Genet.* 254:330-336 (1997), the entirety of which is herein incorporated by reference; Becker *et al.*, *Mol. Gen. Genet.* 249:65-73 (1995), the entirety of which is herein incorporated by reference), potato (Van der Voort *et al.*,



*Mol. Gen. Genet.* 255:438-447 (1997), the entirety of which is herein incorporated by reference; Meksem *et al.*, *Mol. Gen. Genet.* 249:74-81 (1995), the entirety of which is herein incorporated by reference), *Phytophthora infestans* (Van der Lee *et al.*, *Fungal Genet. Biol.* 21:278-291 (1997), the entirety of which is herein incorporated by reference), *Bacillus anthracis* (Keim *et al.*, *J. Bacteriol.* 179:818-824 (1997), the entirety of which is herein incorporated by reference), *Astragalus cremnophylax* (Travis *et al.*, *Mol. Ecol.* 5:735-745 (1996), the entirety of which is herein incorporated by reference), *Arabidopsis* (Cnops *et al.*, *Mol. Gen. Genet.* 253:32-41 (1996), the entirety of which is herein incorporated by reference), *Escherichia coli* (Lin *et al.*, *Nucleic Acids Res.* 24:3649-3650 (1996), the entirety of which is herein incorporated by reference), *Aeromonas* (Huys *et al.*, *Int. J. Syst. Bacteriol.* 46:572-580 (1996), the entirety of which is herein incorporated by reference), nematode (Folkertsma *et al.*, *Mol. Plant Microbe Interact.* 9:47-54 (1996), the entirety of which is herein incorporated by reference), tomato (Thomas *et al.*, *Plant J.* 8:785-794 (1995), the entirety of which is herein incorporated by reference) and human (Latorra *et al.*, *PCR Methods Appl.* 3:351-358 (1994), the entirety of which is herein incorporated by reference). AFLP analysis has also been used for fingerprinting mRNA (Money *et al.*, *Nucleic Acids Res.* 24:2616-2617 (1996), the entirety of which is herein incorporated by reference; Bachem *et al.*, *Plant J.* 9:745-753 (1996), the entirety of which is herein incorporated by reference). It is understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by AFLP analysis or for fingerprinting RNA.

Polymorphisms may also be found using random amplified polymorphic DNA (RAPD) (Williams *et al.*, *Nucl. Acids Res.* 18:6531-6535 (1990), the entirety of which is herein incorporated by reference) and cleaveable amplified polymorphic sequences (CAPS) (Lyamichev

et al., *Science* 260:778-783 (1993), the entirety of which is herein incorporated by reference). It is understood that one or more of the nucleic acid molecules of the present invention, may be utilized as markers or probes to detect polymorphisms by RAPD or CAPS analysis.

Through genetic mapping, a fine scale linkage map can be developed using DNA markers and, then, a genomic DNA library of large-sized fragments can be screened with molecular markers linked to the desired trait. Molecular markers are advantageous for agronomic traits that are otherwise difficult to tag, such as resistance to pathogens, insects and nematodes, tolerance to abiotic stress, quality parameters and quantitative traits such as high yield potential.

The essential requirements for marker-assisted selection in a plant breeding program are: (1) the marker(s) should co-segregate or be closely linked with the desired trait; (2) an efficient means of screening large populations for the molecular marker(s) should be available; and (3) the screening technique should have high reproducibility across laboratories and preferably be economical to use and be user-friendly.

The genetic linkage of marker molecules can be established by a gene mapping model such as, without limitation, the flanking marker model reported by Lander and Botstein, *Genetics* 121:185-199 (1989) and the interval mapping, based on maximum likelihood methods described by Lander and Botstein, *Genetics* 121:185-199 (1989) and implemented in the software package MAPMAKER/QTL (Lincoln and Lander, *Mapping Genes Controlling Quantitative Traits Using MAPMAKER/QTL*, Whitehead Institute for Biomedical Research, Massachusetts, (1990).

Additional software includes Qgene, Version 2.23 (1996), Department of Plant Breeding and Biometry, 266 Emerson Hall, Cornell University, Ithaca, NY, the manual of which is herein incorporated by reference in its entirety). Use of Qgene software is a particularly preferred approach.

A maximum likelihood estimate (MLE) for the presence of a marker is calculated, together with an MLE assuming no QTL effect, to avoid false positives. A  $\log_{10}$  of an odds ratio (LOD) is then calculated as:  $\text{LOD} = \log_{10}(\text{MLE for the presence of a QTL} / \text{MLE given no linked QTL})$ .

The LOD score essentially indicates how much more likely the data are to have arisen assuming the presence of a QTL than in its absence. The LOD threshold value for avoiding a false positive with a given confidence, say 95%, depends on the number of markers and the length of the genome. Graphs indicating LOD thresholds are set forth in Lander and Botstein, *Genetics* 121:185-199 (1989) the entirety of which is herein incorporated by reference and further described by Arús and Moreno-González, *Plant Breeding*, Hayward *et al.*, (eds.) Chapman & Hall, London, pp. 314-331 (1993), the entirety of which is herein incorporated by reference.

Additional models can be used. Many modifications and alternative approaches to interval mapping have been reported, including the use non-parametric methods (Kruglyak and Lander, *Genetics* 139:1421-1428 (1995), the entirety of which is herein incorporated by reference). Multiple regression methods or models can be also be used, in which the trait is regressed on a large number of markers (Jansen, *Biometrics in Plant Breeding*, van Oijen and Jansen (eds.), Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, pp. 116-124 (1994); Weber and Wricke, *Advances in Plant Breeding*, Blackwell, Berlin, 16 (1994), both of which is herein incorporated by reference in their entirety). Procedures combining interval mapping with regression analysis, whereby the phenotype is regressed onto a single putative QTL at a given marker interval and at the same time onto a number of markers that serve as 'cofactors,' have been reported by Jansen and Stam, *Genetics*



136:1447-1455 (1994), the entirety of which is herein incorporated by reference and Zeng, *Genetics* 136:1457-1468 (1994) the entirety of which is herein incorporated by reference. Generally, the use of cofactors reduces the bias and sampling error of the estimated QTL positions (Utz and Melchinger, *Biometrics in Plant Breeding*, van Oijen and Jansen (eds.) Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, pp.195-204 (1994), the entirety of which is herein incorporated by reference, thereby improving the precision and efficiency of QTL mapping (Zeng, *Genetics* 136:1457-1468 (1994)). These models can be extended to multi-environment experiments to analyze genotype-environment interactions (Jansen *et al.*, *Theo. Appl. Genet.* 91:33-37 (1995), the entirety of which is herein incorporated by reference).

Selection of an appropriate mapping populations is important to map construction. The choice of appropriate mapping population depends on the type of marker systems employed (Tanksley *et al.*, *Molecular Mapping Plant Chromosomes. Chromosome Structure and Function: Impact of New Concepts*, Gustafson and Appels (eds.), Plenum Press, New York, pp 157-173 (1988), the entirety of which is herein incorporated by reference). Consideration must be given to the source of parents (adapted vs. exotic) used in the mapping population. Chromosome pairing and recombination rates can be severely disturbed (suppressed) in wide crosses (adapted x exotic) and generally yield greatly reduced linkage distances. Wide crosses will usually provide segregating populations with a relatively large array of polymorphisms when compared to progeny in a narrow cross (adapted x adapted).

An  $F_2$  population is the first generation of selfing after the hybrid seed is produced. Usually a single  $F_1$  plant is selfed to generate a population segregating for all the genes in Mendelian (1:2:1) fashion. Maximum genetic information is obtained from a completely

classified  $F_2$  population using a codominant marker system (Mather, *Measurement of Linkage in Heredity*, Methuen and Co., (1938), the entirety of which is herein incorporated by reference). In the case of dominant markers, progeny tests (e.g.  $F_3$ ,  $BCF_2$ ) are required to identify the heterozygotes, thus making it equivalent to a completely classified  $F_2$  population. However, this procedure is often prohibitive because of the cost and time involved in progeny testing. Progeny testing of  $F_2$  individuals is often used in map construction where phenotypes do not consistently reflect genotype (e.g. disease resistance) or where trait expression is controlled by a QTL. Segregation data from progeny test populations (e.g.  $F_3$  or  $BCF_2$ ) can be used in map construction. Marker-assisted selection can then be applied to cross progeny based on marker-trait map associations ( $F_2$ ,  $F_3$ ), where linkage groups have not been completely disassociated by recombination events (i.e., maximum disequilibrium).

Recombinant inbred lines (RIL) (genetically related lines; usually  $>F_5$ , developed from continuously selfing  $F_2$  lines towards homozygosity) can be used as a mapping population. Information obtained from dominant markers can be maximized by using RIL because all loci are homozygous or nearly so. Under conditions of tight linkage (i.e., about  $<10\%$  recombination), dominant and co-dominant markers evaluated in RIL populations provide more information per individual than either marker type in backcross populations (Reiter *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 89:1477-1481 (1992), the entirety of which is herein incorporated by reference). However, as the distance between markers becomes larger (i.e., loci become more independent), the information in RIL populations decreases dramatically when compared to codominant markers.

Backcross populations (e.g., generated from a cross between a successful variety (recurrent parent) and another variety (donor parent) carrying a trait not present in the former)

can be utilized as a mapping population. A series of backcrosses to the recurrent parent can be made to recover most of its desirable traits. Thus a population is created consisting of individuals nearly like the recurrent parent but each individual carries varying amounts or mosaic of genomic regions from the donor parent. Backcross populations can be useful for mapping dominant markers if all loci in the recurrent parent are homozygous and the donor and recurrent parent have contrasting polymorphic marker alleles (Reiter *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 89:1477-1481 (1992)). Information obtained from backcross populations using either codominant or dominant markers is less than that obtained from  $F_2$  populations because one, rather than two, recombinant gametes are sampled per plant. Backcross populations, however, are more informative (at low marker saturation) when compared to RILs as the distance between linked loci increases in RIL populations (i.e. about 15% recombination). Increased recombination can be beneficial for resolution of tight linkages, but may be undesirable in the construction of maps with low marker saturation.

Near-isogenic lines (NIL) created by many backcrosses to produce an array of individuals that are nearly identical in genetic composition except for the trait or genomic region under interrogation can be used as a mapping population. In mapping with NILs, only a portion of the polymorphic loci are expected to map to a selected region.

Bulk segregant analysis (BSA) is a method developed for the rapid identification of linkage between markers and traits of interest (Michelmore *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 88:9828-9832 (1991), the entirety of which is herein incorporated by reference). In BSA, two bulked DNA samples are drawn from a segregating population originating from a single cross. These bulks contain individuals that are identical for a particular trait (resistant or susceptible to particular disease) or genomic region but arbitrary at unlinked regions (i.e. heterozygous).

Regions unlinked to the target region will not differ between the bulked samples of many individuals in BSA.

It is understood that one or more of the nucleic acid molecules of the present invention may be used as molecular markers. It is also understood that one or more of the protein molecules of the present invention may be used as molecular markers.

In accordance with this aspect of the present invention, a sample nucleic acid is obtained from plants cells or tissues. Any source of nucleic acid may be used. Preferably, the nucleic acid is genomic DNA. The nucleic acid is subjected to restriction endonuclease digestion. For example, one or more nucleic acid molecule or fragment thereof of the present invention can be used as a probe in accordance with the above-described polymorphic methods. The polymorphism obtained in this approach can then be cloned to identify the mutation at the coding region which alters the protein's structure or regulatory region of the gene which affects its expression level.

In an aspect of the present invention, one or more of the nucleic molecules of the present invention are used to determine the level (i.e., the concentration of mRNA in a sample, etc.) in a plant (preferably maize or soybean) or pattern (i.e., the kinetics of expression, rate of decomposition, stability profile, etc.) of the expression of a protein encoded in part or whole by one or more of the nucleic acid molecule of the present invention (collectively, the "Expression Response" of a cell or tissue). As used herein, the Expression Response manifested by a cell or tissue is said to be "altered" if it differs from the Expression Response of cells or tissues of plants not exhibiting the phenotype. To determine whether an Expression Response is altered, the Expression Response manifested by the cell or tissue of the plant exhibiting the phenotype is compared with that of a similar cell or tissue sample of a plant not exhibiting the phenotype. As

will be appreciated, it is not necessary to re-determine the Expression Response of the cell or tissue sample of plants not exhibiting the phenotype each time such a comparison is made; rather, the Expression Response of a particular plant may be compared with previously obtained values of normal plants. As used herein, the phenotype of the organism is any of one or more characteristics of an organism (e.g. disease resistance, pest tolerance, environmental tolerance such as tolerance to abiotic stress, male sterility, quality improvement or yield etc.). A change in genotype or phenotype may be transient or permanent. Also as used herein, a tissue sample is any sample that comprises more than one cell. In a preferred aspect, a tissue sample comprises cells that share a common characteristic (e.g. derived from root, seed, flower, leaf, stem or pollen etc.).

In one aspect of the present invention, an evaluation can be conducted to determine whether a particular mRNA molecule is present. One or more of the nucleic acid molecules of the present invention, preferably one or more of the EST nucleic acid molecules or fragments thereof of the present invention are utilized to detect the presence or quantity of the mRNA species. Such molecules are then incubated with cell or tissue extracts of a plant under conditions sufficient to permit nucleic acid hybridization. The detection of double-stranded probe-mRNA hybrid molecules is indicative of the presence of the mRNA; the amount of such hybrid formed is proportional to the amount of mRNA. Thus, such probes may be used to ascertain the level and extent of the mRNA production in a plant's cells or tissues. Such nucleic acid hybridization may be conducted under quantitative conditions (thereby providing a numerical value of the amount of the mRNA present). Alternatively, the assay may be conducted as a qualitative assay that indicates either that the mRNA is present, or that its level exceeds a user set, predefined value.

A principle of *in situ* hybridization is that a labeled, single-stranded nucleic acid probe will hybridize to a complementary strand of cellular DNA or RNA and, under the appropriate conditions, these molecules will form a stable hybrid. When nucleic acid hybridization is combined with histological techniques, specific DNA or RNA sequences can be identified within a single cell. An advantage of *in situ* hybridization over more conventional techniques for the detection of nucleic acids is that it allows an investigator to determine the precise spatial population (Angerer *et al.*, *Dev. Biol.* 101:477-484 (1984), the entirety of which is herein incorporated by reference; Angerer *et al.*, *Dev. Biol.* 112:157-166 (1985), the entirety of which is herein incorporated by reference; Dixon *et al.*, *EMBO J.* 10:1317-1324 (1991), the entirety of which is herein incorporated by reference). *In situ* hybridization may be used to measure the steady-state level of RNA accumulation. It is a sensitive technique and RNA sequences present in as few as 5-10 copies per cell can be detected (Hardin *et al.*, *J. Mol. Biol.* 202:417-431 (1989), the entirety of which is herein incorporated by reference). A number of protocols have been devised for *in situ* hybridization, each with tissue preparation, hybridization and washing conditions (Meyerowitz, *Plant Mol. Biol. Rep.* 5:242-250 (1987), the entirety of which is herein incorporated by reference; Cox and Goldberg, In: *Plant Molecular Biology: A Practical Approach*, Shaw (ed.), pp 1-35, IRL Press, Oxford (1988), the entirety of which is herein incorporated by reference; Raikhel *et al.*, *In situ RNA hybridization in plant tissues*, In: *Plant Molecular Biology Manual*, vol. B9:1-32, Kluwer Academic Publisher, Dordrecht, Belgium (1989), the entirety of which is herein incorporated by reference).

*In situ* hybridization also allows for the localization of proteins within a tissue or cell (Wilkinson, *In Situ Hybridization*, Oxford University Press, Oxford (1992), the entirety of which is herein incorporated by reference; Langdale, *In Situ Hybridization* In: *The Maize Handbook*,

Freeling and Walbot (eds.), pp 165-179, Springer-Verlag, New York (1994), the entirety of which is herein incorporated by reference). It is understood that one or more of the molecules of the present invention, preferably one or more of the EST nucleic acid molecules or fragments thereof of the present invention or one or more of the antibodies of the present invention may be utilized to detect the level or pattern of a transcription factor or mRNA thereof by *in situ* hybridization.

Fluorescent *in situ* hybridization allows the localization of a particular DNA sequence along a chromosome which is useful, among other uses, for gene mapping, following chromosomes in hybrid lines or detecting chromosomes with translocations, transversions or deletions. *In situ* hybridization has been used to identify chromosomes in several plant species (Griffor *et al.*, *Plant Mol. Biol.* 17:101-109 (1991), the entirety of which is herein incorporated by reference; Gustafson *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 87:1899-1902 (1990), herein incorporated by reference; Mukai and Gill, *Genome* 34:448-452 (1991), the entirety of which is herein incorporated by reference; Schwarzacher and Heslop-Harrison, *Genome* 34:317-323 (1991); Wang *et al.*, *Jpn. J. Genet.* 66:313-316 (1991), the entirety of which is herein incorporated by reference; Parra and Windle, *Nature Genetics* 5:17-21 (1993), the entirety of which is herein incorporated by reference). It is understood that the nucleic acid molecules of the present invention may be used as probes or markers to localize sequences along a chromosome.

Another method to localize the expression of a molecule is tissue printing. Tissue printing provides a way to screen, at the same time on the same membrane many tissue sections from different plants or different developmental stages. Tissue-printing procedures utilize films designed to immobilize proteins and nucleic acids. In essence, a freshly cut section of a tissue is pressed gently onto nitrocellulose paper, nylon membrane or polyvinylidene difluoride

membrane. Such membranes are commercially available (e.g. Millipore, Bedford, Massachusetts U.S.A.). The contents of the cut cell transfer onto the membrane and the contents are immobilized to the membrane. The immobilized contents form a latent print that can be visualized with appropriate probes. When a plant tissue print is made on nitrocellulose paper, the cell walls leave a physical print that makes the anatomy visible without further treatment (Varner and Taylor, *Plant Physiol.* 91:31-33 (1989), the entirety of which is herein incorporated by reference).

Tissue printing on substrate films is described by Daoust, *Exp. Cell Res.* 12:203-211 (1957), the entirety of which is herein incorporated by reference, who detected amylase, protease, ribonuclease and deoxyribonuclease in animal tissues using starch, gelatin and agar films. These techniques can be applied to plant tissues (Yomo and Taylor, *Planta* 112:35-43 (1973); the entirety of which is herein incorporated by reference; Harris and Chrispeels, *Plant Physiol.* 56:292-299 (1975), the entirety of which is herein incorporated by reference). Advances in membrane technology have increased the range of applications of Daoust's tissue-printing techniques allowing (Cassab and Varner, *J. Cell. Biol.* 105:2581-2588 (1987), the entirety of which is herein incorporated by reference) the histochemical localization of various plant enzymes and deoxyribonuclease on nitrocellulose paper and nylon (Spruce *et al.*, *Phytochemistry* 26:2901-2903 (1987), the entirety of which is herein incorporated by reference; Barres *et al.*, *Neuron* 5:527-544 (1990), the entirety of which is herein incorporated by reference; Reid and Pont-Lezica, *Tissue Printing: Tools for the Study of Anatomy, Histochemistry and Gene Expression*, Academic Press, New York, New York (1992), the entirety of which is herein incorporated by reference; Reid *et al.*, *Plant Physiol.* 93:160-165 (1990), the entirety of which is



herein incorporated by reference; Ye *et al.*, *Plant J. 1*:175-183 (1991), the entirety of which is herein incorporated by reference).

It is understood that one or more of the molecules of the present invention, preferably one or more of the EST nucleic acid molecules or fragments thereof of the present invention or one or more of the antibodies of the present invention may be utilized to detect the presence or quantity of a transcription factor by tissue printing.

Further it is also understood that any of the nucleic acid molecules of the present invention may be used as marker nucleic acids and or probes in connection with methods that require probes or marker nucleic acids. As used herein, a probe is an agent that is utilized to determine an attribute or feature (e.g. presence or absence, location, correlation, etc.) of a molecule, cell, tissue or plant. As used herein, a marker nucleic acid is a nucleic acid molecule that is utilized to determine an attribute or feature (e.g., presence or absence, location, correlation, etc.) or a molecule, cell, tissue or plant.

A microarray-based method for high-throughput monitoring of plant gene expression may be utilized to measure gene-specific hybridization targets. This 'chip'-based approach involves using microarrays of nucleic acid molecules as gene-specific hybridization targets to quantitatively measure expression of the corresponding plant genes (Schena *et al.*, *Science* 270:467-470 (1995), the entirety of which is herein incorporated by reference; Shalon, Ph.D. Thesis, Stanford University (1996), the entirety of which is herein incorporated by reference). Every nucleotide in a large sequence can be queried at the same time. Hybridization can be used to efficiently analyze nucleotide sequences.

Several microarray methods have been described. One method compares the sequences to be analyzed by hybridization to a set of oligonucleotides representing all possible

subsequences (Bains and Smith, *J. Theor. Biol.* 135:303-307 (1989), the entirety of which is herein incorporated by reference). A second method hybridizes the sample to an array of oligonucleotide or cDNA molecules. An array consisting of oligonucleotides complementary to subsequences of a target sequence can be used to determine the identity of a target sequence, measure its amount and detect differences between the target and a reference sequence. Nucleic acid molecules microarrays may also be screened with protein molecules or fragments thereof to determine nucleic acid molecules that specifically bind protein molecules or fragments thereof.

The microarray approach may be used with polypeptide targets (U.S. Patent No. 5,445,934; U.S. Patent No: 5,143,854; U.S. Patent No. 5,079,600; U.S. Patent No. 4,923,901, all of which are herein incorporated by reference in their entirety). Essentially, polypeptides are synthesized on a substrate (microarray) and these polypeptides can be screened with either protein molecules or fragments thereof or nucleic acid molecules in order to screen for either protein molecules or fragments thereof or nucleic acid molecules that specifically bind the target polypeptides. (Fodor *et al.*, *Science* 251:767-773 (1991), the entirety of which is herein incorporated by reference). It is understood that one or more of the nucleic acid molecules or protein or fragments thereof of the present invention may be utilized in a microarray based method.

In a preferred embodiment of the present invention microarrays may be prepared that comprise nucleic acid molecules where such nucleic acid molecules encode at least one, preferably at least two, more preferably at least three tetrapyrrole pathway enzymes. In a preferred embodiment the nucleic acid molecules are selected from the group consisting of a putative maize or soybean chlorophyll synthetase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or fragment

thereof, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or fragment thereof, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or fragment thereof, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or fragment thereof and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or fragment thereof.

Site directed mutagenesis may be utilized to modify nucleic acid sequences, particularly as it is a technique that allows one or more of the amino acids encoded by a nucleic acid molecule to be altered (e.g. a threonine to be replaced by a methionine). Three basic methods for site directed mutagenesis are often employed. These are cassette mutagenesis (Wells *et al.*, *Gene* 34:315-323 (1985), the entirety of which is herein incorporated by reference), primer extension (Gilliam *et al.*, *Gene* 12:129-137 (1980), the entirety of which is herein incorporated by reference; Zoller and Smith, *Methods Enzymol.* 100:468-500 (1983), the entirety of which is herein incorporated by reference; Dalbadie-McFarland *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*

79:6409-6413 (1982), the entirety of which is herein incorporated by reference) and methods based upon PCR (Scharf *et al.*, *Science* 233:1076-1078 (1986), the entirety of which is herein incorporated by reference; Higuchi *et al.*, *Nucleic Acids Res.* 16:7351-7367 (1988), the entirety of which is herein incorporated by reference). Site directed mutagenesis approaches are also described in European Patent 0 385 962, the entirety of which is herein incorporated by reference; European Patent 0 359 472, the entirety of which is herein incorporated by reference; and PCT Patent Application WO 93/07278, the entirety of which is herein incorporated by reference.

Site directed mutagenesis strategies have been applied to plants for both *in vitro* as well as *in vivo* site directed mutagenesis (Lanz *et al.*, *J. Biol. Chem.* 266:9971-9976 (1991), the entirety of which is herein incorporated by reference; Kovgan and Zhdanov, *Biotekhnologiya* 5:148-154; No. 207160n, Chemical Abstracts 110:225 (1989), the entirety of which is herein incorporated by reference; Ge *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86:4037-4041 (1989), the entirety of which is herein incorporated by reference; Zhu *et al.*, *J. Biol. Chem.* 271:18494-18498 (1996), the entirety of which is herein incorporated by reference; Chu *et al.*, *Biochemistry* 33:6150-6157 (1994), the entirety of which is herein incorporated by reference; Small *et al.*, *EMBO J.* 11:1291-1296 (1992), the entirety of which is herein incorporated by reference; Cho *et al.*, *Mol. Biotechnol.* 8:13-16 (1997), the entirety of which is herein incorporated by reference; Kita *et al.*, *J. Biol. Chem.* 271:26529-26535 (1996), the entirety of which is herein incorporated by reference; Jin *et al.*, *Mol. Microbiol.* 7:555-562 (1993), the entirety of which is herein incorporated by reference; Hatfield and Vierstra, *J. Biol. Chem.* 267:14799-14803 (1992), the entirety of which is herein incorporated by reference; Zhao *et al.*, *Biochemistry* 31:5093-5099 (1992), the entirety of which is herein incorporated by reference).

Any of the nucleic acid molecules of the present invention may either be modified by site directed mutagenesis or used as, for example, nucleic acid molecules that are used to target other nucleic acid molecules for modification. It is understood that mutants with more than one altered nucleotide can be constructed using techniques that practitioners are familiar with such as isolating restriction fragments and ligating such fragments into an expression vector (*see, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989)).

Sequence-specific DNA-binding proteins play a role in the regulation of transcription. The isolation of recombinant cDNAs encoding these proteins facilitates the biochemical analysis of their structural and functional properties. Genes encoding such DNA-binding proteins have been isolated using classical genetics (Vollbrecht *et al.*, *Nature* 350: 241-243 (1991), the entirety of which is herein incorporated by reference) and molecular biochemical approaches, including the screening of recombinant cDNA libraries with antibodies (Landschulz *et al.*, *Genes Dev.* 2:786-800 (1988), the entirety of which is herein incorporated by reference) or DNA probes (Bodner *et al.*, *Cell* 55:505-518 (1988), the entirety of which is herein incorporated by reference). In addition, an *in situ* screening procedure has been used and has facilitated the isolation of sequence-specific DNA-binding proteins from various plant species (Gilmartin *et al.*, *Plant Cell* 4:839-849 (1992), the entirety of which is herein incorporated by reference; Schindler *et al.*, *EMBO J.* 11:1261-1273 (1992), the entirety of which is herein incorporated by reference). An *in situ* screening protocol does not require the purification of the protein of interest (Vinson *et al.*, *Genes Dev.* 2:801-806 (1988), the entirety of which is herein incorporated by reference; Singh *et al.*, *Cell* 52:415-423 (1988), the entirety of which is herein incorporated by reference).

Two steps may be employed to characterize DNA-protein interactions. The first is to identify promoter fragments that interact with DNA-binding proteins, to titrate binding activity, to determine the specificity of binding and to determine whether a given DNA-binding activity can interact with related DNA sequences (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2<sup>nd</sup> edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)). Electrophoretic mobility-shift assay is a widely used assay. The assay provides a rapid and sensitive method for detecting DNA-binding proteins based on the observation that the mobility of a DNA fragment through a nondenaturing, low-ionic strength polyacrylamide gel is retarded upon association with a DNA-binding protein (Fried and Crother, *Nucleic Acids Res.* 9:6505-6525 (1981), the entirety of which is herein incorporated by reference). When one or more specific binding activities have been identified, the exact sequence of the DNA bound by the protein may be determined. Several procedures for characterizing protein/DNA-binding sites are used, including methylation and ethylation interference assays (Maxam and Gilbert, *Methods Enzymol.* 65:499-560 (1980), the entirety of which is herein incorporated by reference; Wissman and Hillen, *Methods Enzymol.* 208:365-379 (1991), the entirety of which is herein incorporated by reference), footprinting techniques employing DNase I (Galas and Schmitz, *Nucleic Acids Res.* 5:3157-3170 (1978), the entirety of which is herein incorporated by reference), 1,10-phenanthroline-copper ion methods (Sigman *et al.*, *Methods Enzymol.* 208:414-433 (1991), the entirety of which is herein incorporated by reference) and hydroxyl radicals methods (Dixon *et al.*, *Methods Enzymol.* 208:414-433 (1991), the entirety of which is herein incorporated by reference). It is understood that one or more of the nucleic acid molecules of the present invention may be utilized to identify a protein or fragment thereof that specifically binds to a nucleic acid molecule of the present invention. It is also understood that one or more of the

protein molecules or fragments thereof of the present invention may be utilized to identify a nucleic acid molecule that specifically binds to it.

A two-hybrid system is based on the fact that many cellular functions are carried out by proteins, such as transcription factors, that interact (physically) with one another. Two-hybrid systems have been used to probe the function of new proteins (Chien *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 88:9578-9582 (1991) the entirety of which is herein incorporated by reference; Durfee *et al.*, *Genes Dev.* 7:555-569 (1993) the entirety of which is herein incorporated by reference; Choi *et al.*, *Cell* 78:499-512 (1994), the entirety of which is herein incorporated by reference; Kranz *et al.*, *Genes Dev.* 8:313-327 (1994), the entirety of which is herein incorporated by reference).

Interaction mating techniques have facilitated a number of two-hybrid studies of protein-protein interaction. Interaction mating has been used to examine interactions between small sets of tens of proteins (Finley and Brent, *Proc. Natl. Acad. Sci. (U.S.A.)* 91:12098-12984 (1994), the entirety of which is herein incorporated by reference), larger sets of hundreds of proteins (Bendixen *et al.*, *Nucl. Acids Res.* 22:1778-1779 (1994), the entirety of which is herein incorporated by reference) and to comprehensively map proteins encoded by a small genome (Bartel *et al.*, *Nature Genetics* 12:72-77 (1996), the entirety of which is herein incorporated by reference). This technique utilizes proteins fused to the DNA-binding domain and proteins fused to the activation domain. They are expressed in two different haploid yeast strains of opposite mating type and the strains are mated to determine if the two proteins interact. Mating occurs when haploid yeast strains come into contact and result in the fusion of the two haploids into a diploid yeast strain. An interaction can be determined by the activation of a two-hybrid reporter gene in the diploid strain. An advantage of this technique is that it reduces the number of yeast transformations needed to test individual interactions. It is understood that the protein-protein

interactions of protein or fragments thereof of the present invention may be investigated using the two-hybrid system and that any of the nucleic acid molecules of the present invention that encode such proteins or fragments thereof may be used to transform yeast in the two-hybrid system.

**(a) Plant Constructs and Plant Transformants**

One or more of the nucleic acid molecules of the present invention may be used in plant transformation or transfection. Exogenous genetic material may be transferred into a plant cell and the plant cell regenerated into a whole, fertile or sterile plant. Exogenous genetic material is any genetic material, whether naturally occurring or otherwise, from any source that is capable of being inserted into any organism. Such genetic material may be transferred into either monocotyledons and dicotyledons including, but not limited to maize (pp 63-69), soybean (pp 50-60), *Arabidopsis* (p 45), phaseolus (pp 47-49), peanut (pp 49-50), alfalfa (p 60), wheat (pp 69-71), rice (pp 72-79), oat (pp 80-81), sorghum (p 83), rye (p 84), tritordeum (p 84), millet (p85), fescue (p 85), perennial ryegrass (p 86), sugarcane (p87), cranberry (p101), papaya (pp 101-102), banana (p 103), banana (p 103), muskmelon (p 104), apple (p 104), cucumber (p 105), dendrobium (p 109), gladiolus (p 110), chrysanthemum (p 110), liliacea (p 111), cotton (pp113-114), eucalyptus (p 115), sunflower (p 118), canola (p 118), turfgrass (p121), sugarbeet (p 122), coffee (p 122) and dioscorea (p 122), (Christou, In: *Particle Bombardment for Genetic Engineering of Plants*, Biotechnology Intelligence Unit. Academic Press, San Diego, California (1996), the entirety of which is herein incorporated by reference).

Transfer of a nucleic acid that encodes for a protein can result in overexpression of that protein in a transformed cell or transgenic plant. One or more of the proteins or fragments thereof encoded by nucleic acid molecules of the present invention may be overexpressed in a



transformed cell or transformed plant. Particularly, any of the transcription factors or fragments thereof may be overexpressed in a transformed cell or transgenic plant. Such overexpression may be the result of transient or stable transfer of the exogenous genetic material.

Exogenous genetic material may be transferred into a plant cell and the plant cell by the use of a DNA vector or construct designed for such a purpose. Design of such a vector is generally within the skill of the art (See, *Plant Molecular Biology: A Laboratory Manual*, Clark (ed.), Springer, New York (1997), the entirety of which is herein incorporated by reference).

A construct or vector may include a plant promoter to express the protein or protein fragment of choice. A number of promoters which are active in plant cells have been described in the literature. These include the nopaline synthase (NOS) promoter (Ebert *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 84:5745-5749 (1987), the entirety of which is herein incorporated by reference), the octopine synthase (OCS) promoter (which are carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*), the caulimovirus promoters such as the cauliflower mosaic virus (CaMV) 19S promoter (Lawton *et al.*, *Plant Mol. Biol.* 9:315-324 (1987), the entirety of which is herein incorporated by reference) and the CAMV 35S promoter (Odell *et al.*, *Nature* 313:810-812 (1985), the entirety of which is herein incorporated by reference), the figwort mosaic virus 35S-promoter, the light-inducible promoter from the small subunit of ribulose-1,5-bis-phosphate carboxylase (ssRUBISCO), the Adh promoter (Walker *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 84:6624-6628 (1987), the entirety of which is herein incorporated by reference), the sucrose synthase promoter (Yang *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 87:4144-4148 (1990), the entirety of which is herein incorporated by reference), the R gene complex promoter (Chandler *et al.*, *The Plant Cell* 1:1175-1183 (1989), the entirety of which is herein incorporated by reference) and the chlorophyll a/b binding protein gene promoter, etc. These promoters have been used to create

DNA constructs which have been expressed in plants; *see, e.g.*, PCT publication WO 84/02913, herein incorporated by reference in its entirety.

Promoters which are known or are found to cause transcription of DNA in plant cells can be used in the present invention. Such promoters may be obtained from a variety of sources such as plants and plant viruses. It is preferred that the particular promoter selected should be capable of causing sufficient expression to result in the production of an effective amount of the transcription factor to cause the desired phenotype. In addition to promoters that are known to cause transcription of DNA in plant cells, other promoters may be identified for use in the current invention by screening a plant cDNA library for genes which are selectively or preferably expressed in the target tissues or cells.

For the purpose of expression in source tissues of the plant, such as the leaf, seed, root or stem, it is preferred that the promoters utilized in the present invention have relatively high expression in these specific tissues. For this purpose, one may choose from a number of promoters for genes with tissue- or cell-specific or -enhanced expression. Examples of such promoters reported in the literature include the chloroplast glutamine synthetase GS2 promoter from pea (Edwards *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 87:3459-3463 (1990), herein incorporated by reference in its entirety), the chloroplast fructose-1,6-biphosphatase (FBPase) promoter from wheat (Lloyd *et al.*, *Mol. Gen. Genet.* 225:209-216 (1991), herein incorporated by reference in its entirety), the nuclear photosynthetic ST-LS1 promoter from potato (Stockhaus *et al.*, *EMBO J.* 8:2445-2451 (1989), herein incorporated by reference in its entirety), the serine/threonine kinase (PAL) promoter and the glucoamylase (CHS) promoter from *Arabidopsis thaliana*. Also reported to be active in photosynthetically active tissues are the ribulose-1,5-bisphosphate carboxylase (RbcS) promoter from eastern larch (*Larix laricina*), the promoter for

the *cab* gene, *cab6*, from pine (Yamamoto *et al.*, *Plant Cell Physiol.* 35:773-778 (1994), herein incorporated by reference in its entirety), the promoter for the *Cab-1* gene from wheat (Fejes *et al.*, *Plant Mol. Biol.* 15:921-932 (1990), herein incorporated by reference in its entirety), the promoter for the *CAB-1* gene from spinach (Lubberstedt *et al.*, *Plant Physiol.* 104:997-1006 (1994), herein incorporated by reference in its entirety), the promoter for the *cab1R* gene from rice (Luan *et al.*, *Plant Cell.* 4:971-981 (1992), the entirety of which is herein incorporated by reference), the pyruvate, orthophosphate dikinase (PPDK) promoter from maize (Matsuoka *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 90: 9586-9590 (1993), herein incorporated by reference in its entirety), the promoter for the tobacco *Lhcb1\*2* gene (Cerdan *et al.*, *Plant Mol. Biol.* 33:245-255 (1997), herein incorporated by reference in its entirety), the *Arabidopsis thaliana* *SUC2* sucrose-H<sup>+</sup> symporter promoter (Truernit *et al.*, *Planta.* 196:564-570 (1995), herein incorporated by reference in its entirety) and the promoter for the thylakoid membrane proteins from spinach (*psaD*, *psaF*, *psaE*, *PC*, *FNR*, *atpC*, *atpD*, *cab*, *rbcS*). Other promoters for the chlorophyll a/b-binding proteins may also be utilized in the present invention, such as the promoters for *Lhcb* gene and *PsbP* gene from white mustard (*Sinapis alba*; Kretsch *et al.*, *Plant Mol. Biol.* 28:219-229 (1995), the entirety of which is herein incorporated by reference).

For the purpose of expression in sink tissues of the plant, such as the tuber of the potato plant, the fruit of tomato, or the seed of maize, wheat, rice and barley, it is preferred that the promoters utilized in the present invention have relatively high expression in these specific tissues. A number of promoters for genes with tuber-specific or -enhanced expression are known, including the class I patatin promoter (Bevan *et al.*, *EMBO J.* 8:1899-1906 (1986); Jefferson *et al.*, *Plant Mol. Biol.* 14:995-1006 (1990), both of which are herein incorporated by reference in its entirety), the promoter for the potato tuber ADPGPP genes, both the large and

small subunits, the sucrose synthase promoter (Salanoubat and Belliard, *Gene*. 60:47-56 (1987), Salanoubat and Belliard, *Gene*. 84:181-185 (1989), both of which are incorporated by reference in their entirety), the promoter for the major tuber proteins including the 22 kd protein complexes and proteinase inhibitors (Hannapel, *Plant Physiol*. 101:703-704 (1993), herein incorporated by reference in its entirety), the promoter for the granule bound starch synthase gene (GBSS) (Visser *et al.*, *Plant Mol. Biol.* 17:691-699 (1991), herein incorporated by reference in its entirety) and other class I and II patatins promoters (Koster-Topfer *et al.*, *Mol Gen Genet*. 219:390-396 (1989); Mignery *et al.*, *Gene*. 62:27-44 (1988), both of which are herein incorporated by reference in their entirety).

Other promoters can also be used to express a transcription factor or fragment thereof in specific tissues, such as seeds or fruits. The promoter for  $\beta$ -conglycinin (Chen *et al.*, *Dev. Genet*. 10: 112-122 (1989), herein incorporated by reference in its entirety) or other seed-specific promoters such as the napin and phaseolin promoters, can be used. The zeins are a group of storage proteins found in maize endosperm. Genomic clones for zein genes have been isolated (Pedersen *et al.*, *Cell* 29:1015-1026 (1982), herein incorporated by reference in its entirety) and the promoters from these clones, including the 15 kD, 16 kD, 19 kD, 22 kD, 27 kD and  $\gamma$  genes, could also be used. Other promoters known to function, for example, in maize include the promoters for the following genes: *waxy*, *Brittle*, *Shrunken 2*, Branching enzymes I and II, starch synthases, debranching enzymes, oleosins, glutelins and sucrose synthases. A particularly preferred promoter for maize endosperm expression is the promoter for the glutelin gene from rice, more particularly the Osgt-1 promoter (Zheng *et al.*, *Mol. Cell Biol*. 13:5829-5842 (1993), herein incorporated by reference in its entirety). Examples of promoters suitable for expression

in wheat include those promoters for the ADPglucose pyrosynthase (ADPGPP) subunits, the granule bound and other starch synthase, the branching and debranching enzymes, the embryogenesis-abundant proteins, the gliadins and the glutenins. Examples of such promoters in rice include those promoters for the ADPGPP subunits, the granule bound and other starch synthase, the branching enzymes, the debranching enzymes, sucrose synthases and the glutelins. A particularly preferred promoter is the promoter for rice glutelin, Osgt-1. Examples of such promoters for barley include those for the ADPGPP subunits, the granule bound and other starch synthase, the branching enzymes, the debranching enzymes, sucrose synthases, the hordeins, the embryo globulins and the aleurone specific proteins.

Root specific promoters may also be used. An example of such a promoter is the promoter for the acid chitinase gene (Samac *et al.*, *Plant Mol. Biol.* 25:587-596 (1994), the entirety of which is herein incorporated by reference). Expression in root tissue could also be accomplished by utilizing the root specific subdomains of the CaMV35S promoter that have been identified (Lam *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86:7890-7894 (1989), herein incorporated by reference in its entirety). Other root cell specific promoters include those reported by Conkling *et al.* (Conkling *et al.*, *Plant Physiol.* 93:1203-1211 (1990), the entirety of which is herein incorporated by reference).

Additional promoters that may be utilized are described, for example, in U.S. Patent Nos. 5,378,619; 5,391,725; 5,428,147; 5,447,858; 5,608,144; 5,608,144; 5,614,399; 5,633,441; 5,633,435; and 4,633,436, all of which are herein incorporated in their entirety. In addition, a tissue specific enhancer may be used (Fromm *et al.*, *The Plant Cell* 1:977-984 (1989), the entirety of which is herein incorporated by reference).

Constructs or vectors may also include with the coding region of interest a nucleic acid sequence that acts, in whole or in part, to terminate transcription of that region. For example, such sequences have been isolated including the Tr7 3' sequence and the NOS 3' sequence (Ingelbrecht *et al.*, *The Plant Cell* 1:671-680 (1989), the entirety of which is herein incorporated by reference; Bevan *et al.*, *Nucleic Acids Res.* 11:369-385 (1983), the entirety of which is herein incorporated by reference), or the like.

A vector or construct may also include regulatory elements. Examples of such include the Adh intron 1 (Callis *et al.*, *Genes and Develop.* 1:1183-1200 (1987), the entirety of which is herein incorporated by reference), the sucrose synthase intron (Vasil *et al.*, *Plant Physiol.* 91:1575-1579 (1989), the entirety of which is herein incorporated by reference) and the TMV omega element (Gallie *et al.*, *The Plant Cell* 1:301-311 (1989), the entirety of which is herein incorporated by reference). These and other regulatory elements may be included when appropriate.

A vector or construct may also include a selectable marker. Selectable markers may also be used to select for plants or plant cells that contain the exogenous genetic material. Examples of such include, but are not limited to, a neo gene (Potrykus *et al.*, *Mol. Gen. Genet.* 199:183-188 (1985), the entirety of which is herein incorporated by reference) which codes for kanamycin resistance and can be selected for using kanamycin, G418, etc.; a bar gene which codes for bialaphos resistance; a mutant EPSP synthase gene (Hinchey *et al.*, *Bio/Technology* 6:915-922 (1988), the entirety of which is herein incorporated by reference) which encodes glyphosate resistance; a nitrilase gene which confers resistance to bromoxynil (Stalker *et al.*, *J. Biol. Chem.* 263:6310-6314 (1988), the entirety of which is herein incorporated by reference); a mutant acetolactate synthase gene (ALS) which confers imidazolinone or sulphonylurea resistance

(European Patent Application 154,204 (Sept. 11, 1985), the entirety of which is herein incorporated by reference); and a methotrexate resistant DHFR gene (Thillet *et al.*, *J. Biol. Chem.* 263:12500-12508 (1988), the entirety of which is herein incorporated by reference).

A vector or construct may also include a transit peptide. Incorporation of a suitable chloroplast transit peptide may also be employed (European Patent Application Publication Number 0218571, the entirety of which is herein incorporated by reference). Translational enhancers may also be incorporated as part of the vector DNA. DNA constructs could contain one or more 5' non-translated leader sequences which may serve to enhance expression of the gene products from the resulting mRNA transcripts. Such sequences may be derived from the promoter selected to express the gene or can be specifically modified to increase translation of the mRNA. Such regions may also be obtained from viral RNAs, from suitable eukaryotic genes, or from a synthetic gene sequence. For a review of optimizing expression of transgenes, see Koziel *et al.*, *Plant Mol. Biol.* 32:393-405 (1996), the entirety of which is herein incorporated by reference.

A vector or construct may also include a screenable marker. Screenable markers may be used to monitor expression. Exemplary screenable markers include a  $\beta$ -glucuronidase or uidA gene (GUS) which encodes an enzyme for which various chromogenic substrates are known (Jefferson, *Plant Mol. Biol. Rep.* 5:387-405 (1987), the entirety of which is herein incorporated by reference; Jefferson *et al.*, *EMBO J.* 6:3901-3907 (1987), the entirety of which is herein incorporated by reference); an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta *et al.*, *Stadler Symposium* 11:263-282 (1988), the entirety of which is herein incorporated by reference); a  $\beta$ -

lactamase gene (Sutcliffe *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 75:3737-3741 (1978), the entirety of which is herein incorporated by reference), a gene which encodes an enzyme for which various chromogenic substrates are known (e.g., PADAC, a chromogenic cephalosporin); a luciferase gene (Ow *et al.*, *Science* 234:856-859 (1986), the entirety of which is herein incorporated by reference); a xyle gene (Zukowsky *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 80:1101-1105 (1983), the entirety of which is herein incorporated by reference) which encodes a catechol dioxygenase that can convert chromogenic catechols; an  $\alpha$ -amylase gene (Ikata *et al.*, *Bio/Technol.* 8:241-242 (1990), the entirety of which is herein incorporated by reference); a tyrosinase gene (Katz *et al.*, *J. Gen. Microbiol.* 129:2703-2714 (1983), the entirety of which is herein incorporated by reference) which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to melanin; an  $\alpha$ -galactosidase, which will turn a chromogenic  $\alpha$ -galactose substrate.

Included within the terms "selectable or screenable marker genes" are also genes which encode a secretable marker whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers which encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes which can be detected catalytically. Secretable proteins fall into a number of classes, including small, diffusible proteins which are detectable, (e.g., by ELISA), small active enzymes which are detectable in extracellular solution (e.g.,  $\alpha$ -amylase,  $\beta$ -lactamase, phosphinothricin transferase), or proteins which are inserted or trapped in the cell wall (such as proteins which include a leader sequence such as that found in the expression unit of extension or tobacco PR-S). Other possible selectable and/or screenable marker genes will be apparent to those of skill in the art.



There are many methods for introducing transforming nucleic acid molecules into plant cells. Suitable methods are believed to include virtually any method by which nucleic acid molecules may be introduced into a cell, such as by *Agrobacterium* infection or direct delivery of nucleic acid molecules such as, for example, by PEG-mediated transformation, by electroporation or by acceleration of DNA coated particles, etc (Potrykus, *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 42:205-225 (1991), the entirety of which is herein incorporated by reference; Vasil, *Plant Mol. Biol.* 25:925-937 (1994), the entirety of which is herein incorporated by reference). For example, electroporation has been used to transform maize protoplasts (Fromm *et al.*, *Nature* 312:791-793 (1986), the entirety of which is herein incorporated by reference).

Other vector systems suitable for introducing transforming DNA into a host plant cell include but are not limited to binary artificial chromosome (BIBAC) vectors (Hamilton *et al.*, *Gene* 200:107-116 (1997), the entirety of which is herein incorporated by reference); and transfection with RNA viral vectors (Della-Cioppa *et al.*, *Ann. N.Y. Acad. Sci.* (1996), 792 (Engineering Plants for Commercial Products and Applications), 57-61, the entirety of which is herein incorporated by reference). Additional vector systems also include plant selectable YAC vectors such as those described in Mullen *et al.*, *Molecular Breeding* 4:449-457 (1988), the entirety of which is herein incorporated by reference).

Technology for introduction of DNA into cells is well known to those of skill in the art. Four general methods for delivering a gene into cells have been described: (1) chemical methods (Graham and van der Eb, *Virology* 54:536-539 (1973), the entirety of which is herein incorporated by reference); (2) physical methods such as microinjection (Capecchi, *Cell* 22:479-488 (1980), the entirety of which is herein incorporated by reference), electroporation (Wong and

Neumann, *Biochem. Biophys. Res. Commun.* 107:584-587 (1982); Fromm *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 82:5824-5828 (1985); U.S. Patent No. 5,384,253, all of which are herein incorporated in their entirety); and the gene gun (Johnston and Tang, *Methods Cell Biol.* 43:353-365 (1994), the entirety of which is herein incorporated by reference); (3) viral vectors (Clapp, *Clin. Perinatol.* 20:155-168 (1993); Lu *et al.*, *J. Exp. Med.* 178:2089-2096 (1993); Eglitis and Anderson, *Biotechniques* 6:608-614 (1988), all of which are herein incorporated in their entirety); and (4) receptor-mediated mechanisms (Curiel *et al.*, *Hum. Gen. Ther.* 3:147-154 (1992), Wagner *et al.*, *Proc. Natl. Acad. Sci. (USA)* 89:6099-6103 (1992), both of which are incorporated by reference in their entirety).

Acceleration methods that may be used include, for example, microprojectile bombardment and the like. One example of a method for delivering transforming nucleic acid molecules to plant cells is microprojectile bombardment. This method has been reviewed by Yang and Christou (eds.), *Particle Bombardment Technology for Gene Transfer*, Oxford Press, Oxford, England (1994), the entirety of which is herein incorporated by reference). Non-biological particles (microprojectiles) that may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum and the like.

A particular advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly transforming monocots, is that neither the isolation of protoplasts (Cristou *et al.*, *Plant Physiol.* 87:671-674 (1988), the entirety of which is herein incorporated by reference) nor the susceptibility of *Agrobacterium* infection are required. An illustrative embodiment of a method for delivering DNA into maize cells by acceleration is a

biolistics  $\alpha$ -particle delivery system, which can be used to propel particles coated with DNA through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with corn cells cultured in suspension. Gordon-Kamm *et al.*, describes the basic procedure for coating tungsten particles with DNA (Gordon-Kamm *et al.*, *Plant Cell* 2:603-618 (1990), the entirety of which is herein incorporated by reference). The screen disperses the tungsten nucleic acid particles so that they are not delivered to the recipient cells in large aggregates. A particle delivery system suitable for use with the present invention is the helium acceleration PDS-1000/He gun is available from Bio-Rad Laboratories (Bio-Rad, Hercules, California)(Sanford *et al.*, *Technique* 3:3-16 (1991), the entirety of which is herein incorporated by reference).

For the bombardment, cells in suspension may be concentrated on filters. Filters containing the cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate. If desired, one or more screens are also positioned between the gun and the cells to be bombarded.

Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate. If desired, one or more screens are also positioned between the acceleration device and the cells to be bombarded. Through the use of techniques set forth herein one may obtain up to 1000 or more foci of cells transiently expressing a marker gene. The number of cells in a focus which express the exogenous gene product 48 hours post-bombardment often range from one to ten and average one to three.

In bombardment transformation, one may optimize the pre-bombardment culturing conditions and the bombardment parameters to yield the maximum numbers of stable

transformants. Both the physical and biological parameters for bombardment are important in this technology. Physical factors are those that involve manipulating the DNA/microprojectile precipitate or those that affect the flight and velocity of either the macro- or microprojectiles. Biological factors include all steps involved in manipulation of cells before and immediately after bombardment, the osmotic adjustment of target cells to help alleviate the trauma associated with bombardment and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. It is believed that pre-bombardment manipulations are especially important for successful transformation of immature embryos.

In another alternative embodiment, plastids can be stably transformed. Methods disclosed for plastid transformation in higher plants include the particle gun delivery of DNA containing a selectable marker and targeting of the DNA to the plastid genome through homologous recombination (Svab *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 87:8526-8530 (1990); Svab and Maliga, *Proc. Natl. Acad. Sci. (U.S.A.)* 90:913-917 (1993); Staub and Maliga, *EMBO J.* 12:601-606 (1993); U.S. Patent Nos. 5, 451,513 and 5,545,818, all of which are herein incorporated by reference in their entirety).

Accordingly, it is contemplated that one may wish to adjust various aspects of the bombardment parameters in small scale studies to fully optimize the conditions. One may particularly wish to adjust physical parameters such as gap distance, flight distance, tissue distance and helium pressure. One may also minimize the trauma reduction factors by modifying conditions which influence the physiological state of the recipient cells and which may therefore influence transformation and integration efficiencies. For example, the osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells may be adjusted for

optimum transformation. The execution of other routine adjustments will be known to those of skill in the art in light of the present disclosure.

*Agrobacterium*-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for example the methods described by Fraley *et al.*, *Bio/Technology* 3:629-635 (1985) and Rogers *et al.*, *Methods Enzymol.* 153:253-277 (1987), both of which are herein incorporated by reference in their entirety. Further, the integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences and intervening DNA is usually inserted into the plant genome as described (Spielmann *et al.*, *Mol. Gen. Genet.* 205:34 (1986), the entirety of which is herein incorporated by reference).

Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as described (Klee *et al.*, In: *Plant DNA Infectious Agents*, Hohn and Schell (eds.), Springer-Verlag, New York, pp. 179-203 (1985), the entirety of which is herein incorporated by reference. Moreover, technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes (Rogers *et al.*, *Methods Enzymol.* 153:253-277 (1987)). In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used

for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

A transgenic plant formed using *Agrobacterium* transformation methods typically contains a single gene on one chromosome. Such transgenic plants can be referred to as being heterozygous for the added gene. More preferred is a transgenic plant that is homozygous for the added structural gene; *i.e.*, a transgenic plant that contains two added genes, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) an independent segregant transgenic plant that contains a single added gene, germinating some of the seed produced and analyzing the resulting plants produced for the gene of interest.

It is also to be understood that two different transgenic plants can also be mated to produce offspring that contain two independently segregating added, exogenous genes. Selfing of appropriate progeny can produce plants that are homozygous for both added, exogenous genes that encode a polypeptide of interest. Back-crossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated, as is vegetative propagation.

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation and combinations of these treatments (*See, for example*, Potrykus *et al.*, *Mol. Gen. Genet.* 205:193-200 (1986); Lorz *et al.*, *Mol. Gen. Genet.* 199:178 (1985); Fromm *et al.*, *Nature* 319:791 (1986); Uchimiya *et al.*, *Mol. Gen. Genet.* 204:204 (1986); Marcotte *et al.*, *Nature* 335:454-457 (1988), all of which are herein incorporated by reference in their entirety).

Application of these systems to different plant strains depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration

of cereals from protoplasts are described (Fujimura *et al.*, *Plant Tissue Culture Letters* 2:74 (1985); Toriyama *et al.*, *Theor Appl. Genet.* 205:34 (1986); Yamada *et al.*, *Plant Cell Rep.* 4:85 (1986); Abdullah *et al.*, *Biotechnology* 4:1087 (1986), all of which are herein incorporated by reference in their entirety).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil, *Biotechnology* 6:397 (1988), the entirety of which is herein incorporated by reference). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil *et al.*, *Bio/Technology* 10:667 (1992), the entirety of which is herein incorporated by reference).

Using the latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al.*, *Nature* 328:70 (1987); Klein *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 85:8502-8505 (1988); McCabe *et al.*, *Bio/Technology* 6:923 (1988), all of which are herein incorporated by reference in their entirety). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants.

Other methods of cell transformation can also be used and include but are not limited to introduction of DNA into plants by direct DNA transfer into pollen (Zhou *et al.*, *Methods Enzymol.* 101:433 (1983); Hess *et al.*, *Intern Rev. Cytol.* 107:367 (1987); Luo *et al.*, *Plant Mol Biol. Reporter* 6:165 (1988), all of which are herein incorporated by reference in their entirety), by direct injection of DNA into reproductive organs of a plant (Pena *et al.*, *Nature* 325:274 (1987), the entirety of which is herein incorporated by reference), or by direct injection of DNA into the cells of immature embryos followed by the rehydration of desiccated embryos (Neuhaus

et al., *Theor. Appl. Genet.* 75:30 (1987), the entirety of which is herein incorporated by reference).

The regeneration, development and cultivation of plants from single plant protoplast transformants or from various transformed explants is well known in the art (Weissbach and Weissbach, In: *Methods for Plant Molecular Biology*, Academic Press, San Diego, CA, (1988), the entirety of which is herein incorporated by reference). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil.

The development or regeneration of plants containing the foreign, exogenous gene that encodes a protein of interest is well known in the art. Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important lines. Conversely, pollen from plants of these important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

There are a variety of methods for the regeneration of plants from plant tissue. The particular method of regeneration will depend on the starting plant tissue and the particular plant species to be regenerated.

Methods for transforming dicots, primarily by use of *Agrobacterium tumefaciens* and obtaining transgenic plants have been published for cotton (U.S. Patent No. 5,004,863; U.S. Patent No. 5,159,135; U.S. Patent No. 5,518,908, all of which are herein incorporated by



reference in their entirety); soybean (U.S. Patent No. 5,569,834; U.S. Patent No. 5,416,011; McCabe *et al.*, *Biotechnology* 6:923 (1988); Christou *et al.*, *Plant Physiol.* 87:671-674 (1988); all of which are herein incorporated by reference in their entirety); *Brassica* (U.S. Patent No. 5,463,174, the entirety of which is herein incorporated by reference); peanut (Cheng *et al.*, *Plant Cell Rep.* 15:653-657 (1996), McKently *et al.*, *Plant Cell Rep.* 14:699-703 (1995), all of which are herein incorporated by reference in their entirety); papaya; and pea (Grant *et al.*, *Plant Cell Rep.* 15:254-258 (1995), the entirety of which is herein incorporated by reference).

Transformation of monocotyledons using electroporation, particle bombardment and *Agrobacterium* have also been reported. Transformation and plant regeneration have been achieved in asparagus (Bytebier *et al.*, *Proc. Natl. Acad. Sci. (USA)* 84:5354 (1987), the entirety of which is herein incorporated by reference); barley (Wan and Lemaux, *Plant Physiol* 104:37 (1994), the entirety of which is herein incorporated by reference); maize (Rhodes *et al.*, *Science* 240:204 (1988); Gordon-Kamm *et al.*, *Plant Cell* 2:603-618 (1990); Fromm *et al.*, *Bio/Technology* 8:833 (1990); Koziel *et al.*, *Bio/Technology* 11:194 (1993); Armstrong *et al.*, *Crop Science* 35:550-557 (1995); all of which are herein incorporated by reference in their entirety); oat (Somers *et al.*, *Bio/Technology* 10:1589 (1992), the entirety of which is herein incorporated by reference); orchard grass (Horn *et al.*, *Plant Cell Rep.* 7:469 (1988), the entirety of which is herein incorporated by reference); rice (Toriyama *et al.*, *Theor Appl. Genet.* 205:34 (1986); Part *et al.*, *Plant Mol. Biol.* 32:1135-1148 (1996); Abedinia *et al.*, *Aust. J. Plant Physiol.* 24:133-141 (1997); Zhang and Wu, *Theor. Appl. Genet.* 76:835 (1988); Zhang *et al.*, *Plant Cell Rep.* 7:379 (1988); Battraw and Hall, *Plant Sci.* 86:191-202 (1992); Christou *et al.*, *Bio/Technology* 9:957 (1991), all of which are herein incorporated by reference in their entirety); rye (De la Pena *et al.*, *Nature* 325:274 (1987), the entirety of which is herein incorporated by

reference); sugarcane (Bower and Birch, *Plant J.* 2:409 (1992), the entirety of which is herein incorporated by reference); tall fescue (Wang *et al.*, *Bio/Technology* 10:691 (1992), the entirety of which is herein incorporated by reference) and wheat (Vasil *et al.*, *Bio/Technology* 10:667 (1992), the entirety of which is herein incorporated by reference; U.S. Patent No. 5,631,152, the entirety of which is herein incorporated by reference.)

Assays for gene expression based on the transient expression of cloned nucleic acid constructs have been developed by introducing the nucleic acid molecules into plant cells by polyethylene glycol treatment, electroporation, or particle bombardment (Marcotte *et al.*, *Nature* 335:454-457 (1988), the entirety of which is herein incorporated by reference; Marcotte *et al.*, *Plant Cell* 1:523-532 (1989), the entirety of which is herein incorporated by reference; McCarty *et al.*, *Cell* 66:895-905 (1991), the entirety of which is herein incorporated by reference; Hattori *et al.*, *Genes Dev.* 6:609-618 (1992), the entirety of which is herein incorporated by reference; Goff *et al.*, *EMBO J.* 9:2517-2522 (1990), the entirety of which is herein incorporated by reference). Transient expression systems may be used to functionally dissect gene constructs (*see generally*, Mailga *et al.*, *Methods in Plant Molecular Biology*, Cold Spring Harbor Press (1995)).

Any of the nucleic acid molecules of the present invention may be introduced into a plant cell in a permanent or transient manner in combination with other genetic elements such as vectors, promoters, enhancers etc. Further, any of the nucleic acid molecules of the present invention may be introduced into a plant cell in a manner that allows for overexpression of the protein or fragment thereof encoded by the nucleic acid molecule.

Cosuppression is the reduction in expression levels, usually at the level of RNA, of a particular endogenous gene or gene family by the expression of a homologous sense construct

that is capable of transcribing mRNA of the same strandedness as the transcript of the endogenous gene (Napoli *et al.*, *Plant Cell* 2:279-289 (1990), the entirety of which is herein incorporated by reference; van der Krol *et al.*, *Plant Cell* 2:291-299 (1990), the entirety of which is herein incorporated by reference). Cosuppression may result from stable transformation with a single copy nucleic acid molecule that is homologous to a nucleic acid sequence found with the cell (Prolls and Meyer, *Plant J.* 2:465-475 (1992), the entirety of which is herein incorporated by reference) or with multiple copies of a nucleic acid molecule that is homologous to a nucleic acid sequence found with the cell (Mittlesten *et al.*, *Mol. Gen. Genet.* 244:325-330 (1994), the entirety of which is herein incorporated by reference). Genes, even though different, linked to homologous promoters may result in the cosuppression of the linked genes (Vaucheret, *C.R. Acad. Sci. III* 316:1471-1483 (1993), the entirety of which is herein incorporated by reference).

This technique has, for example, been applied to generate white flowers from red petunia and tomatoes that do not ripen on the vine. Up to 50% of petunia transformants that contained a sense copy of the glucoamylase (CHS) gene produced white flowers or floral sectors; this was as a result of the post-transcriptional loss of mRNA encoding CHS (Flavell, *Proc. Natl. Acad. Sci. (U.S.A.)* 91:3490-3496 (1994), the entirety of which is herein incorporated by reference); van Blokland *et al.*, *Plant J.* 6:861-877 (1994), the entirety of which is herein incorporated by reference). Cosuppression may require the coordinate transcription of the transgene and the endogenous gene and can be reset by a developmental control mechanism (Jorgensen, *Trends Biotechnol.* 8:340-344 (1990), the entirety of which is herein incorporated by reference; Meins and Kunz, In: *Gene Inactivation and Homologous Recombination in Plants*, Paszkowski (ed.), pp. 335-348, Kluwer Academic, Netherlands (1994), the entirety of which is herein incorporated by reference).

It is understood that one or more of the nucleic acids of the present invention may be introduced into a plant cell and transcribed using an appropriate promoter with such transcription resulting in the cosuppression of an endogenous transcription factor.

Antisense approaches are a way of preventing or reducing gene function by targeting the genetic material (Mol *et al.*, *FEBS Lett.* 268:427-430 (1990), the entirety of which is herein incorporated by reference). The objective of the antisense approach is to use a sequence complementary to the target gene to block its expression and create a mutant cell line or organism in which the level of a single chosen protein is selectively reduced or abolished. Antisense techniques have several advantages over other 'reverse genetic' approaches. The site of inactivation and its developmental effect can be manipulated by the choice of promoter for antisense genes or by the timing of external application or microinjection. Antisense can manipulate its specificity by selecting either unique regions of the target gene or regions where it shares homology to other related genes (Hiatt *et al.*, In: *Genetic Engineering*, Setlow (ed.), Vol. 11, New York: Plenum 49-63 (1989), the entirety of which is herein incorporated by reference).

The principle of regulation by antisense RNA is that RNA that is complementary to the target mRNA is introduced into cells, resulting in specific RNA:RNA duplexes being formed by base pairing between the antisense substrate and the target mRNA (Green *et al.*, *Annu. Rev. Biochem.* 55:569-597 (1986), the entirety of which is herein incorporated by reference). Under one embodiment, the process involves the introduction and expression of an antisense gene sequence. Such a sequence is one in which part or all of the normal gene sequences are placed under a promoter in inverted orientation so that the 'wrong' or complementary strand is transcribed into a noncoding antisense RNA that hybridizes with the target mRNA and interferes with its expression (Takayama and Inouye, *Crit. Rev. Biochem. Mol. Biol.* 25:155-184 (1990),

the entirety of which is herein incorporated by reference). An antisense vector is constructed by standard procedures and introduced into cells by transformation, transfection, electroporation, microinjection, infection, etc. The type of transformation and choice of vector will determine whether expression is transient or stable. The promoter used for the antisense gene may influence the level, timing, tissue, specificity, or inducibility of the antisense inhibition.

It is understood that the activity of a transcription factor in a plant cell may be reduced or depressed by growing a transformed plant cell containing a nucleic acid molecule whose non-transcribed strand encodes a transcription factor or fragment thereof.

Antibodies have been expressed in plants (Hiatt *et al.*, *Nature* 342:76-78 (1989), the entirety of which is herein incorporated by reference; Conrad and Fielder, *Plant Mol. Biol.* 26:1023-1030 (1994), the entirety of which is herein incorporated by reference). Cytoplasmic expression of a scFv (single-chain Fv antibodies) has been reported to delay infection by artichoke mottled crinkle virus. Transgenic plants that express antibodies directed against endogenous proteins may exhibit a physiological effect (Philips *et al.*, *EMBO J.* 16:4489-4496 (1997), the entirety of which is herein incorporated by reference; Marion-Poll, *Trends in Plant Science* 2:447-448 (1997), the entirety of which is herein incorporated by reference). For example, expressed anti-abscisic antibodies have been reported to result in a general perturbation of seed development (Philips *et al.*, *EMBO J.* 16: 4489-4496 (1997)).

Antibodies that are catalytic may also be expressed in plants (abzymes). The principle behind abzymes is that since antibodies may be raised against many molecules, this recognition ability can be directed toward generating antibodies that bind transition states to force a chemical reaction forward (Persidas, *Nature Biotechnology* 15:1313-1315 (1997), the entirety of which is herein incorporated by reference; Baca *et al.*, *Ann. Rev. Biophys. Biomol. Struct.* 26:461-493

(1997), the entirety of which is herein incorporated by reference). The catalytic abilities of abzymes may be enhanced by site directed mutagenesis. Examples of abzymes are, for example, set forth in U.S. Patent No. 5,658,753; U.S. Patent No. 5,632,990; U.S. Patent No. 5,631,137; U.S. Patent No. 5,602,015; U.S. Patent No. 5,559,538; U.S. Patent No. 5,576,174; U.S. Patent No. 5,500,358; U.S. Patent No. 5,318,897; U.S. Patent No. 5,298,409; U.S. Patent No. 5,258,289 and U.S. Patent No. 5,194,585, all of which are herein incorporated in their entirety.

It is understood that any of the antibodies of the present invention may be expressed in plants and that such expression can result in a physiological effect. It is also understood that any of the expressed antibodies may be catalytic.

#### **(b) Fungal Constructs and Fungal Transformants**

The present invention also relates to a fungal recombinant vector comprising exogenous genetic material. The present invention also relates to a fungal cell comprising a fungal recombinant vector. The present invention also relates to methods for obtaining a recombinant fungal host cell comprising introducing into a fungal host cell exogenous genetic material.

Exogenous genetic material may be transferred into a fungal cell. In a preferred embodiment the exogenous genetic material includes a nucleic acid molecule of the present invention having a sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof or fragments of either or other nucleic acid molecule of the present invention. The fungal recombinant vector may be any vector which can be conveniently subjected to recombinant DNA procedures. The choice of a vector will typically depend on the compatibility of the vector with the fungal host cell into which the vector is to be introduced. The vector may be a linear or a closed circular plasmid. The vector system may be a single

vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the fungal host.

The fungal vector may be an autonomously replicating vector, *i.e.*, a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the fungal cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. For integration, the vector may rely on the nucleic acid sequence of the vector for stable integration of the vector into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the fungal host. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, there should be preferably two nucleic acid sequences which individually contain a sufficient number of nucleic acids, preferably 400bp to 1500bp, more preferably 800bp to 1000bp, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. These nucleic acid sequences may be any sequence that is homologous with a target sequence in the genome of the fungal host cell and, furthermore, may be non-encoding or encoding sequences.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of origin of replications for use in a yeast host cell are the 2 micron origin of replication and the combination

of CEN3 and ARS 1. Any origin of replication may be used which is compatible with the fungal host cell of choice.

The fungal vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides, for example biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs and the like. The selectable marker may be selected from the group including, but not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferase), *bar* (phosphinothricin acetyltransferase), *hygB* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase) and *sC* (sulfate adenylyltransferase) and *trpC* (anthranilate synthase). Preferred for use in an *Aspergillus* cell are the *amdS* and *pyrG* markers of *Aspergillus nidulans* or *Aspergillus oryzae* and the *bar* marker of *Streptomyces hygroscopicus*. Furthermore, selection may be accomplished by co-transformation, e.g., as described in WO 91/17243, the entirety of which is herein incorporated by reference. A nucleic acid sequence of the present invention may be operably linked to a suitable promoter sequence. The promoter sequence is a nucleic acid sequence which is recognized by the fungal host cell for expression of the nucleic acid sequence. The promoter sequence contains transcription and translation control sequences which mediate the expression of the protein or fragment thereof.

A promoter may be any nucleic acid sequence which shows transcriptional activity in the fungal host cell of choice and may be obtained from genes encoding polypeptides either homologous or heterologous to the host cell. Examples of suitable promoters for directing the transcription of a nucleic acid construct of the invention in a filamentous fungal host are promoters obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase; *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable



alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase and hybrids thereof. In a yeast host, a useful promoter is the *Saccharomyces cerevisiae* enolase (eno-1) promoter. Particularly preferred promoters are the TAKA amylase, NA2-tpi (a hybrid of the promoters from the genes encoding *Aspergillus niger* neutral alpha -amylase and *Aspergillus oryzae* triose phosphate isomerase) and glaA promoters.

A protein or fragment thereof encoding nucleic acid molecule of the present invention may also be operably linked to a terminator sequence at its 3' terminus. The terminator sequence may be native to the nucleic acid sequence encoding the protein or fragment thereof or may be obtained from foreign sources. Any terminator which is functional in the fungal host cell of choice may be used in the present invention, but particularly preferred terminators are obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase and *Saccharomyces cerevisiae* enolase.

A protein or fragment thereof encoding nucleic acid molecule of the present invention may also be operably linked to a suitable leader sequence. A leader sequence is a nontranslated region of a mRNA which is important for translation by the fungal host. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the protein or fragment thereof. The leader sequence may be native to the nucleic acid sequence encoding the protein or fragment thereof or may be obtained from foreign sources. Any leader sequence which is functional in the fungal host cell of choice may be used in the present invention, but particularly

preferred leaders are obtained from the genes encoding *Aspergillus oryzae* TKA amylase and *Aspergillus oryzae* triose phosphate isomerase.

A polyadenylation sequence may also be operably linked to the 3' terminus of the nucleic acid sequence of the present invention. The polyadenylation sequence is a sequence which when transcribed is recognized by the fungal host to add polyadenosine residues to transcribed mRNA. The polyadenylation sequence may be native to the nucleic acid sequence encoding the protein or fragment thereof or may be obtained from foreign sources. Any polyadenylation sequence which is functional in the fungal host of choice may be used in the present invention, but particularly preferred polyadenylation sequences are obtained from the genes encoding *Aspergillus oryzae* TKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase and *Aspergillus niger* alpha-glucosidase.

To avoid the necessity of disrupting the cell to obtain the protein or fragment thereof and to minimize the amount of possible degradation of the expressed protein or fragment thereof within the cell, it is preferred that expression of the protein or fragment thereof gives rise to a product secreted outside the cell. To this end, a protein or fragment thereof of the present invention may be linked to a signal peptide linked to the amino terminus of the protein or fragment thereof. A signal peptide is an amino acid sequence which permits the secretion of the protein or fragment thereof from the fungal host into the culture medium. The signal peptide may be native to the protein or fragment thereof of the invention or may be obtained from foreign sources. The 5' end of the coding sequence of the nucleic acid sequence of the present invention may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted protein or fragment thereof. Alternatively, the 5' end of the coding sequence may contain a signal peptide

coding region which is foreign to that portion of the coding sequence which encodes the secreted protein or fragment thereof. The foreign signal peptide may be required where the coding sequence does not normally contain a signal peptide coding region. Alternatively, the foreign signal peptide may simply replace the natural signal peptide to obtain enhanced secretion of the desired protein or fragment thereof. The foreign signal peptide coding region may be obtained from a glucoamylase or an amylase gene from an *Aspergillus* species, a lipase or proteinase gene from *Rhizomucor miehei*, the gene for the alpha-factor from *Saccharomyces cerevisiae*, or the calf preprochymosin gene. An effective signal peptide for fungal host cells is the *Aspergillus oryzae* TAKA amylase signal, *Aspergillus niger* neutral amylase signal, the *Rhizomucor miehei* aspartic proteinase signal, the *Humicola lanuginosus* cellulase signal, or the *Rhizomucor miehei* lipase signal. However, any signal peptide capable of permitting secretion of the protein or fragment thereof in a fungal host of choice may be used in the present invention.

A protein or fragment thereof encoding nucleic acid molecule of the present invention may also be linked to a propeptide coding region. A propeptide is an amino acid sequence found at the amino terminus of a protein or proenzyme. Cleavage of the propeptide from the proprotein yields a mature biochemically active protein. The resulting polypeptide is known as a propolypeptide or proenzyme (or a zymogen in some cases). Propolypeptides are generally inactive and can be converted to mature active polypeptides by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide or proenzyme. The propeptide coding region may be native to the protein or fragment thereof or may be obtained from foreign sources. The foreign propeptide coding region may be obtained from the *Saccharomyces cerevisiae* alpha-factor gene or *Myceliophthora thermophila* laccase gene (WO 95/33836, the entirety of which is herein incorporated by reference).

The procedures used to ligate the elements described above to construct the recombinant expression vector of the present invention are well known to one skilled in the art (see, for example, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor, N.Y., (1989)).

The present invention also relates to recombinant fungal host cells produced by the methods of the present invention which are advantageously used with the recombinant vector of the present invention. The cell is preferably transformed with a vector comprising a nucleic acid sequence of the invention followed by integration of the vector into the host chromosome. The choice of fungal host cells will to a large extent depend upon the gene encoding the protein or fragment thereof and its source. The fungal host cell may, for example, be a yeast cell or a filamentous fungal cell.

"Yeast" as used herein includes *Ascosporogenous* yeast (*Endomycetales*), *Basidiosporogenous* yeast and yeast belonging to the *Fungi Imperfecti* (*Blastomycetes*). The *Ascosporogenous* yeasts are divided into the families *Spermophthoraceae* and *Saccharomycetaceae*. The latter is comprised of four subfamilies, *Schizosaccharomycoideae* (for example, genus *Schizosaccharomyces*), *Nadsonioideae*, *Lipomycoideae* and *Saccharomycoideae* (for example, genera *Pichia*, *Kluyveromyces* and *Saccharomyces*). The *Basidiosporogenous* yeasts include the genera *Leucosporidim*, *Rhodosporidium*, *Sporidiobolus*, *Filobasidium* and *Filobasidiella*. Yeast belonging to the *Fungi Imperfecti* are divided into two families, *Sporobolomycetaceae* (for example, genera *Sorobolomyces* and *Bullera*) and *Cryptococcaceae* (for example, genus *Candida*). Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in *Biology and Activities of Yeast* (Skinner *et al.*, *Soc. App. Bacteriol. Symposium Series* No. 9, (1980), the entirety of which is

herein incorporated by reference). The biology of yeast and manipulation of yeast genetics are well known in the art (see, for example, *Biochemistry and Genetics of Yeast*, Bacil *et al.* (ed.), 2nd edition, 1987; *The Yeasts*, Rose and Harrison (eds.), 2nd ed., (1987); and *The Molecular Biology of the Yeast Saccharomyces*, Strathern *et al.* (eds.), (1981), all of which are herein incorporated by reference in their entirety).

"Fungi" as used herein includes the phyla *Ascomycota*, *Basidiomycota*, *Chytridiomycota* and *Zygomycota* (as defined by Hawksworth *et al.*, In: Ainsworth and Bisby's *Dictionary of The Fungi*, 8th edition, 1995, CAB International, University Press, Cambridge, UK; the entirety of which is herein incorporated by reference) as well as the *Oomycota* (as cited in Hawksworth *et al.*, In: Ainsworth and Bisby's *Dictionary of The Fungi*, 8th edition, 1995, CAB International, University Press, Cambridge, UK) and all mitosporic fungi (Hawksworth *et al.*, In: Ainsworth and Bisby's *Dictionary of The Fungi*, 8th edition, 1995, CAB International, University Press, Cambridge, UK). Representative groups of *Ascomycota* include, for example, *Neurospora*, *Eupenicillium* (= *Penicillium*), *Emericella* (= *Aspergillus*), *Eurotium* (= *Aspergillus*) and the true yeasts listed above. Examples of *Basidiomycota* include mushrooms, rusts and smuts. Representative groups of *Chytridiomycota* include, for example, *Allomyces*, *Blastocladiella*, *Coelomomyces* and aquatic fungi. Representative groups of *Oomycota* include, for example, *Saprolegniomycetous* aquatic fungi (water molds) such as *Achlya*. Examples of mitosporic fungi include *Aspergillus*, *Penicillium*, *Candida* and *Alternaria*. Representative groups of *Zygomycota* include, for example, *Rhizopus* and *Mucor*.

"Filamentous fungi" include all filamentous forms of the subdivision *Eumycota* and *Oomycota* (as defined by Hawksworth *et al.*, In: Ainsworth and Bisby's *Dictionary of The Fungi*, 8th edition, 1995, CAB International, University Press, Cambridge, UK). The filamentous fungi

are characterized by a vegetative mycelium composed of chitin, cellulose, glucan, chitosan, mannan and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a unicellular thallus and carbon catabolism may be fermentative.

In one embodiment, the fungal host cell is a yeast cell. In a preferred embodiment, the yeast host cell is a cell of the species of *Candida*, *Kluyveromyces*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia* and *Yarrowia*. In a preferred embodiment, the yeast host cell is a *Saccharomyces cerevisiae* cell, a *Saccharomyces carlsbergensis*, *Saccharomyces diastaticus* cell, a *Saccharomyces douglasii* cell, a *Saccharomyces kluyveri* cell, a *Saccharomyces norbensis* cell, or a *Saccharomyces oviformis* cell. In another preferred embodiment, the yeast host cell is a *Kluyveromyces lactis* cell. In another preferred embodiment, the yeast host cell is a *Yarrowia lipolytica* cell.

In another embodiment, the fungal host cell is a filamentous fungal cell. In a preferred embodiment, the filamentous fungal host cell is a cell of the species of, but not limited to, *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Myceliophthora*, *Mucor*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium* and *Trichoderma*. In a preferred embodiment, the filamentous fungal host cell is an *Aspergillus* cell. In another preferred embodiment, the filamentous fungal host cell is an *Acremonium* cell. In another preferred embodiment, the filamentous fungal host cell is a *Fusarium* cell. In another preferred embodiment, the filamentous fungal host cell is a *Humicola* cell. In another preferred embodiment, the filamentous fungal host cell is a *Myceliophthora* cell. In another even preferred embodiment, the filamentous fungal host cell is a *Mucor* cell. In another preferred embodiment, the filamentous

fungal host cell is a *Neurospora* cell. In another preferred embodiment, the filamentous fungal host cell is a *Penicillium* cell. In another preferred embodiment, the filamentous fungal host cell is a *Thielavia* cell. In another preferred embodiment, the filamentous fungal host cell is a *Tolypocladium* cell. In another preferred embodiment, the filamentous fungal host cell is a *Trichoderma* cell. In a preferred embodiment, the filamentous fungal host cell is an *Aspergillus oryzae* cell, an *Aspergillus niger* cell, an *Aspergillus foetidus* cell, or an *Aspergillus japonicus* cell. In another preferred embodiment, the filamentous fungal host cell is a *Fusarium oxysporum* cell or a *Fusarium graminearum* cell. In another preferred embodiment, the filamentous fungal host cell is a *Humicola insolens* cell or a *Humicola lanuginosus* cell. In another preferred embodiment, the filamentous fungal host cell is a *Myceliophthora thermophila* cell. In a most preferred embodiment, the filamentous fungal host cell is a *Mucor miehei* cell. In a most preferred embodiment, the filamentous fungal host cell is a *Neurospora crassa* cell. In a most preferred embodiment, the filamentous fungal host cell is a *Penicillium purpurogenum* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Thielavia terrestris* cell. In another most preferred embodiment, the *Trichoderma* cell is a *Trichoderma reesei* cell, a *Trichoderma viride* cell, a *Trichoderma longibrachiatum* cell, a *Trichoderma harzianum* cell, or a *Trichoderma koningii* cell. In a preferred embodiment, the fungal host cell is selected from an *A. nidulans* cell, an *A. niger* cell, an *A. oryzae* cell and an *A. sojae* cell. In a further preferred embodiment, the fungal host cell is an *A. nidulans* cell.

The recombinant fungal host cells of the present invention may further comprise one or more sequences which encode one or more factors that are advantageous in the expression of the protein or fragment thereof, for example, an activator (e.g., a trans-acting factor), a chaperone and a processing protease. The nucleic acids encoding one or more of these factors are

preferably not operably linked to the nucleic acid encoding the protein or fragment thereof. An activator is a protein which activates transcription of a nucleic acid sequence encoding a polypeptide (Kudla *et al.*, *EMBO* 9:1355-1364(1990); Jarai and Buxton, *Current Genetics* 26:2238-244(1994); Verdier, *Yeast* 6:271-297(1990), all of which are herein incorporated by reference in their entirety). The nucleic acid sequence encoding an activator may be obtained from the genes encoding *Saccharomyces cerevisiae* heme activator protein 1 (hap1), *Saccharomyces cerevisiae* galactose metabolizing protein 4 (gal4) and *Aspergillus nidulans* ammonia regulation protein (areA). For further examples, see Verdier, *Yeast* 6:271-297 (1990); MacKenzie *et al.*, *Journal of Gen. Microbiol.* 139:2295-2307 (1993), both of which are herein incorporated by reference in their entirety). A chaperone is a protein which assists another protein in folding properly (Hartl *et al.*, *TIBS* 19:20-25 (1994); Bergeron *et al.*, *TIBS* 19:124-128 (1994); Demolder *et al.*, *J. Biotechnology* 32:179-189 (1994); Craig, *Science* 260:1902-1903(1993); Gething and Sambrook, *Nature* 355:33-45 (1992); Puig and Gilbert, *J Biol. Chem.* 269:7764-7771 (1994); Wang and Tsou, *FASEB Journal* 7:1515-11157 (1993); Robinson *et al.*, *Bio/Technology* 1:381-384 (1994), all of which are herein incorporated by reference in their entirety). The nucleic acid sequence encoding a chaperone may be obtained from the genes encoding *Aspergillus oryzae* protein disulphide isomerase, *Saccharomyces cerevisiae* calnexin, *Saccharomyces cerevisiae* BiP/GRP78 and *Saccharomyces cerevisiae* Hsp70. For further examples, see Gething and Sambrook, *Nature* 355:33-45 (1992); Hartl *et al.*, *TIBS* 19:20-25 (1994). A processing protease is a protease that cleaves a propeptide to generate a mature biochemically active polypeptide (Enderlin and Ogrydziak, *Yeast* 10:67-79 (1994); Fuller *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86:1434-1438 (1989); Julius *et al.*, *Cell* 37:1075-1089 (1984); Julius *et al.*, *Cell* 32:839-852 (1983), all of which are incorporated by reference in their entirety).



The nucleic acid sequence encoding a processing protease may be obtained from the genes encoding *Aspergillus niger* Kex2, *Saccharomyces cerevisiae* dipeptidylaminopeptidase, *Saccharomyces cerevisiae* Kex2 and *Yarrowia lipolytica* dibasic processing endoprotease (xpr6). Any factor that is functional in the fungal host cell of choice may be used in the present invention.

Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts and regeneration of the cell wall in a manner known per se. Suitable procedures for transformation of *Aspergillus* host cells are described in EP 238 023 and Yelton *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 81:1470-1474 (1984), both of which are herein incorporated by reference in their entirety. A suitable method of transforming *Fusarium* species is described by Malardier *et al.*, *Gene* 78:147-156 (1989), the entirety of which is herein incorporated by reference. Yeast may be transformed using the procedures described by Becker and Guarente, In: Abelson and Simon, (eds.), *Guide to Yeast Genetics and Molecular Biology, Methods Enzymol.* Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, *J. Bacteriology* 153:163 (1983); Hinnen *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 75:1920 (1978), all of which are herein incorporated by reference in their entirety.

The present invention also relates to methods of producing the protein or fragment thereof comprising culturing the recombinant fungal host cells under conditions conducive for expression of the protein or fragment thereof. The fungal cells of the present invention are cultivated in a nutrient medium suitable for production of the protein or fragment thereof using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under

conditions allowing the protein or fragment thereof to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art (*see, e.g.*, Bennett and LaSure (eds.), *More Gene Manipulations in Fungi*, Academic Press, CA, (1991), the entirety of which is herein incorporated by reference). Suitable media are available from commercial suppliers or may be prepared according to published compositions (*e.g.*, in catalogues of the American Type Culture Collection, Manassas, VA). If the protein or fragment thereof is secreted into the nutrient medium, a protein or fragment thereof can be recovered directly from the medium. If the protein or fragment thereof is not secreted, it is recovered from cell lysates.

The expressed protein or fragment thereof may be detected using methods known in the art that are specific for the particular protein or fragment. These detection methods may include the use of specific antibodies, formation of an enzyme product, or disappearance of an enzyme substrate. For example, if the protein or fragment thereof has enzymatic activity, an enzyme assay may be used. Alternatively, if polyclonal or monoclonal antibodies specific to the protein or fragment thereof are available, immunoassays may be employed using the antibodies to the protein or fragment thereof. The techniques of enzyme assay and immunoassay are well known to those skilled in the art.

The resulting protein or fragment thereof may be recovered by methods known in the arts. For example, the protein or fragment thereof may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation. The recovered protein or fragment thereof may then be further purified by a variety of chromatographic procedures, *e.g.*, ion exchange chromatography, gel filtration chromatography, affinity chromatography, or the like.

**(c) Mammalian Constructs and Transformed Mammalian Cells**

The present invention also relates to methods for obtaining a recombinant mammalian host cell, comprising introducing into a mammalian host cell exogenous genetic material. The present invention also relates to a mammalian cell comprising a mammalian recombinant vector. The present invention also relates to methods for obtaining a recombinant mammalian host cell, comprising introducing into a mammalian cell exogenous genetic material. In a preferred embodiment the exogenous genetic material includes a nucleic acid molecule of the present invention having a sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof or fragments of either or other nucleic acid molecule of the present invention.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC, Manassas, VA), such as HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers *et al.*, *Nature* 273:113 (1978), the entirety of which is herein incorporated by reference), Rous sarcoma virus (RSV), adenovirus (ADV) and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly-A addition sequences. Enhancer sequences which increase expression may also be included and sequences which promote amplification of the gene may also be desirable (for example methotrexate resistance genes).

Vectors suitable for replication in mammalian cells may include viral replicons, or sequences which insure integration of the appropriate sequences encoding HCV epitopes into the host genome. For example, another vector used to express foreign DNA is vaccinia virus. In this

case, for example, a nucleic acid molecule encoding a protein or fragment thereof is inserted into the vaccinia genome. Techniques for the insertion of foreign DNA into the vaccinia virus genome are known in the art and may utilize, for example, homologous recombination. Such heterologous DNA is generally inserted into a gene which is non-essential to the virus, for example, the thymidine kinase gene (tk), which also provides a selectable marker. Plasmid vectors that greatly facilitate the construction of recombinant viruses have been described (*see*, for example, Mackett *et al*, *J Virol.* 49:857 (1984); Chakrabarti *et al.*, *Mol. Cell. Biol.* 5:3403 (1985); Moss, In: *Gene Transfer Vectors For Mammalian Cells* (Miller and Calos, eds., Cold Spring Harbor Laboratory, N.Y., p. 10, (1987); all of which are herein incorporated by reference in their entirety). Expression of the HCV polypeptide then occurs in cells or animals which are infected with the live recombinant vaccinia virus.

The sequence to be integrated into the mammalian sequence may be introduced into the primary host by any convenient means, which includes calcium precipitated DNA, spheroplast fusion, transformation, electroporation, biolistics, lipofection, microinjection, or other convenient means. Where an amplifiable gene is being employed, the amplifiable gene may serve as the selection marker for selecting hosts into which the amplifiable gene has been introduced. Alternatively, one may include with the amplifiable gene another marker, such as a drug resistance marker, e.g. neomycin resistance (G418 in mammalian cells), hygromycin in resistance etc., or an auxotrophy marker (HIS3, TRP1, LEU2, URA3, ADE2, LYS2, etc.) for use in yeast cells.

Depending upon the nature of the modification and associated targeting construct, various techniques may be employed for identifying targeted integration. Conveniently, the DNA may be digested with one or more restriction enzymes and the fragments probed with an appropriate

DNA fragment which will identify the properly sized restriction fragment associated with integration.

One may use different promoter sequences, enhancer sequences, or other sequence which will allow for enhanced levels of expression in the expression host. Thus, one may combine an enhancer from one source, a promoter region from another source, a 5'- noncoding region upstream from the initiation methionine from the same or different source as the other sequences and the like. One may provide for an intron in the non-coding region with appropriate splice sites or for an alternative 3'- untranslated sequence or polyadenylation site. Depending upon the particular purpose of the modification, any of these sequences may be introduced, as desired.

Where selection is intended, the sequence to be integrated will have with it a marker gene, which allows for selection. The marker gene may conveniently be downstream from the target gene and may include resistance to a cytotoxic agent, e.g. antibiotics, heavy metals, or the like, resistance or susceptibility to HAT, gancyclovir, etc., complementation to an auxotrophic host, particularly by using an auxotrophic yeast as the host for the subject manipulations, or the like. The marker gene may also be on a separate DNA molecule, particularly with primary mammalian cells. Alternatively, one may screen the various transformants, due to the high efficiency of recombination in yeast, by using hybridization analysis, PCR, sequencing, or the like.

For homologous recombination, constructs can be prepared where the amplifiable gene will be flanked, normally on both sides with DNA homologous with the DNA of the target region. Depending upon the nature of the integrating DNA and the purpose of the integration, the homologous DNA will generally be within 100kb, usually 50kb, preferably about 25kb, of the transcribed region of the target gene, more preferably within 2kb of the target gene. Where

modeling of the gene is intended, homology will usually be present proximal to the site of the mutation. The homologous DNA may include the 5'-upstream region outside of the transcriptional regulatory region or comprising any enhancer sequences, transcriptional initiation sequences, adjacent sequences, or the like. The homologous region may include a portion of the coding region, where the coding region may be comprised only of an open reading frame or combination of exons and introns. The homologous region may comprise all or a portion of an intron, where all or a portion of one or more exons may also be present. Alternatively, the homologous region may comprise the 3'-region, so as to comprise all or a portion of the transcriptional termination region, or the region 3' of this region. The homologous regions may extend over all or a portion of the target gene or be outside the target gene comprising all or a portion of the transcriptional regulatory regions and/or the structural gene.

The integrating constructs may be prepared in accordance with conventional ways, where sequences may be synthesized, isolated from natural sources, manipulated, cloned, ligated, subjected to in vitro mutagenesis, primer repair, or the like. At various stages, the joined sequences may be cloned and analyzed by restriction analysis, sequencing, or the like. Usually during the preparation of a construct where various fragments are joined, the fragments, intermediate constructs and constructs will be carried on a cloning vector comprising a replication system functional in a prokaryotic host, e.g., *E. coli* and a marker for selection, e.g., biocide resistance, complementation to an auxotrophic host, etc. Other functional sequences may also be present, such as polylinkers, for ease of introduction and excision of the construct or portions thereof, or the like. A large number of cloning vectors are available such as pBR322, the pUC series, etc. These constructs may then be used for integration into the primary mammalian host.

In the case of the primary mammalian host, a replicating vector may be used. Usually, such vector will have a viral replication system, such as SV40, bovine papilloma virus, adenovirus, or the like. The linear DNA sequence vector may also have a selectable marker for identifying transfected cells. Selectable markers include the neo gene, allowing for selection with G418, the herpes tk gene for selection with HAT medium, the gpt gene with mycophenolic acid, complementation of an auxotrophic host, etc.

The vector may or may not be capable of stable maintenance in the host. Where the vector is capable of stable maintenance, the cells will be screened for homologous integration of the vector into the genome of the host, where various techniques for curing the cells may be employed. Where the vector is not capable of stable maintenance, for example, where a temperature sensitive replication system is employed, one may change the temperature from the permissive temperature to the non-permissive temperature, so that the cells may be cured of the vector. In this case, only those cells having integration of the construct comprising the amplifiable gene and, when present, the selectable marker, will be able to survive selection.

Where a selectable marker is present, one may select for the presence of the targeting construct by means of the selectable marker. Where the selectable marker is not present, one may select for the presence of the construct by the amplifiable gene. For the neo gene or the herpes tk gene, one could employ a medium for growth of the transformants of about 0.1-1 mg/ml of G418 or may use HAT medium, respectively. Where DHFR is the amplifiable gene, the selective medium may include from about 0.01-0.5  $\mu$ M of methotrexate or be deficient in glycine-hypoxanthine-thymidine and have dialysed serum (GHT media).

The DNA can be introduced into the expression host by a variety of techniques that include calcium phosphate/DNA co-precipitates, microinjection of DNA into the nucleus,

electroporation, yeast protoplast fusion with intact cells, transfection, polycations, e.g., polybrene, polyornithine, etc., or the like. The DNA may be single or double stranded DNA, linear or circular. The various techniques for transforming mammalian cells are well known (see Keown *et al.*, *Methods Enzymol.* (1989); Keown *et al.*, *Methods Enzymol.* 185:527-537 (1990); Mansour *et al.*, *Nature* 336:348-352, (1988); all of which are herein incorporated by reference in their entirety).

**(d) Insect Constructs and Transformed Insect Cells**

The present invention also relates to an insect recombinant vectors comprising exogenous genetic material. The present invention also relates to an insect cell comprising an insect recombinant vector. The present invention also relates to methods for obtaining a recombinant insect host cell, comprising introducing into an insect cell exogenous genetic material. In a preferred embodiment the exogenous genetic material includes a nucleic acid molecule of the present invention having a sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof or fragments of either or other nucleic acid molecule of the present invention.

The insect recombinant vector may be any vector which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the nucleic acid sequence. The choice of a vector will typically depend on the compatibility of the vector with the insect host cell into which the vector is to be introduced. The vector may be a linear or a closed circular plasmid. The vector system may be a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the insect host. In addition, the insect vector may be an expression vector. Nucleic acid molecules can be suitably inserted into a replication vector for expression in the insect cell under a suitable



promoter for insect cells. Many vectors are available for this purpose and selection of the appropriate vector will depend mainly on the size of the nucleic acid molecule to be inserted into the vector and the particular host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the particular host cell with which it is compatible. The vector components for insect cell transformation generally include, but are not limited to, one or more of the following: a signal sequence, origin of replication, one or more marker genes and an inducible promoter.

The insect vector may be an autonomously replicating vector, *i.e.*, a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the insect cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. For integration, the vector may rely on the nucleic acid sequence of the vector for stable integration of the vector into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the insect host. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, there should be preferably two nucleic acid sequences which individually contain a sufficient number of nucleic acids, preferably 400bp to 1500bp, more preferably 800bp to 1000bp, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. These nucleic acid sequences may be any sequence that is homologous with a

target sequence in the genome of the insect host cell and, furthermore, may be non-encoding or encoding sequences.

Baculovirus expression vectors (BEVs) have become important tools for the expression of foreign genes, both for basic research and for the production of proteins with direct clinical applications in human and veterinary medicine (Doerfler, *Curr. Top. Microbiol. Immunol.* 131:51-68 (1968); Luckow and Summers, *Bio/Technology* 6:47-55 (1988a); Miller, *Annual Review of Microbiol.* 42:177-199 (1988); Summers, *Curr. Comm. Molecular Biology*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1988); all of which are herein incorporated by reference in their entirety). BEVs are recombinant insect viruses in which the coding sequence for a chosen foreign gene has been inserted behind a baculovirus promoter in place of the viral gene, e.g., polyhedrin (Smith and Summers, U.S. Pat. No., 4,745,051, the entirety of which is incorporated herein by reference).

The use of baculovirus vectors relies upon the host cells being derived from *Lepidopteran* insects such as *Spodoptera frugiperda* or *Trichoplusia ni*. The preferred *Spodoptera frugiperda* cell line is the cell line Sf9. The *Spodoptera frugiperda* Sf9 cell line was obtained from American Type Culture Collection (Manassas, VA.) and is assigned accession number ATCC CRL 1711 (Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Ag. Exper. Station Bulletin No. 1555 (1988), the entirety of which is herein incorporated by reference). Other insect cell systems, such as the silkworm *B. mori* may also be used.

The proteins expressed by the BEVs are, therefore, synthesized, modified and transported in host cells derived from *Lepidopteran* insects. Most of the genes that have been inserted and produced in the baculovirus expression vector system have been derived from vertebrate species.

Other baculovirus genes in addition to the polyhedrin promoter may be employed to advantage in a baculovirus expression system. These include immediate-early ( $\alpha$ ), delayed-early ( $\beta$ ), late ( $\gamma$ ), or very late ( $\delta$ ), according to the phase of the viral infection during which they are expressed. The expression of these genes occurs sequentially, probably as the result of a "cascade" mechanism of transcriptional regulation. (Guarino and Summers, *J. Virol.* 57:563-571 (1986); Guarino and Summers, *J. Virol.* 61:2091-2099 (1987); Guarino and Summers, *Virol.* 162:444-451 (1988); all of which are herein incorporated by reference in their entirety).

Insect recombinant vectors are useful as intermediates for the infection or transformation of insect cell systems. For example, an insect recombinant vector containing a nucleic acid molecule encoding a baculovirus transcriptional promoter followed downstream by an insect signal DNA sequence is capable of directing the secretion of the desired biologically active protein from the insect cell. The vector may utilize a baculovirus transcriptional promoter region derived from any of the over 500 baculoviruses generally infecting insects, such as for example the Orders *Lepidoptera*, *Diptera*, *Orthoptera*, *Coleoptera* and *Hymenoptera*, including for example but not limited to the viral DNAs of *Autographa californica* MNPV, *Bombyx mori* NPV, *Trichoplusia ni* MNPV, *Rachiplusia ou* MNPV or *Galleria mellonella* MNPV, wherein said baculovirus transcriptional promoter is a baculovirus immediate-early gene IEl or IEN promoter; an immediate-early gene in combination with a baculovirus delayed-early gene promoter region selected from the group consisting of 39K and a *HindIII-k* fragment delayed-early gene; or a baculovirus late gene promoter. The immediate-early or delayed-early promoters can be enhanced with transcriptional enhancer elements. The insect signal DNA sequence may code for a signal peptide of a *Lepidopteran* adipokinetic hormone precursor or a signal peptide of the *Manduca sexta* adipokinetic hormone precursor (Summers, U.S. Patent No. 5,155,037; the

entirety of which is herein incorporated by reference). Other insect signal DNA sequences include a signal peptide of the *Orthoptera Schistocerca gregaria* locust adipokinetic hormone precursor and the *Drosophila melanogaster* cuticle genes CP1, CP2, CP3 or CP4 or for an insect signal peptide having substantially a similar chemical composition and function (Summers, U.S. Patent No. 5,155,037).

Insect cells are distinctly different from animal cells. Insects have a unique life cycle and have distinct cellular properties such as the lack of intracellular plasminogen activators in which are present in vertebrate cells. Another difference is the high expression levels of protein products ranging from 1 to greater than 500 mg/liter and the ease at which cDNA can be cloned into cells (Frasier, *In Vitro Cell. Dev. Biol.* 25:225 (1989); Summers and Smith, In: *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Ag. Exper. Station Bulletin No. 1555 (1988), both of which are incorporated by reference in their entirety).

Recombinant protein expression in insect cells is achieved by viral infection or stable transformation. For viral infection, the desired gene is cloned into baculovirus at the site of the wild-type polyhedron gene (Webb and Summers, *Technique* 2:173 (1990); Bishop and Posse, *Adv. Gene Technol.* 1:55 (1990); both of which are incorporated by reference in their entirety). The polyhedron gene is a component of a protein coat in occlusions which encapsulate virus particles. Deletion or insertion in the polyhedron gene results the failure to form occlusion bodies. Occlusion negative viruses are morphologically different from occlusion positive viruses and enable one skilled in the art to identify and purify recombinant viruses.

The vectors of present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides, for example biocide or viral resistance, resistance to heavy metals, prototrophy to

auxotrophs and the like. Selection may be accomplished by co-transformation, *e.g.*, as described in WO 91/17243, a nucleic acid sequence of the present invention may be operably linked to a suitable promoter sequence. The promoter sequence is a nucleic acid sequence which is recognized by the insect host cell for expression of the nucleic acid sequence. The promoter sequence contains transcription and translation control sequences which mediate the expression of the protein or fragment thereof. The promoter may be any nucleic acid sequence which shows transcriptional activity in the insect host cell of choice and may be obtained from genes encoding polypeptides either homologous or heterologous to the host cell.

For example, a nucleic acid molecule encoding a protein or fragment thereof may also be operably linked to a suitable leader sequence. A leader sequence is a nontranslated region of a mRNA which is important for translation by the fungal host. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the protein or fragment thereof. The leader sequence may be native to the nucleic acid sequence encoding the protein or fragment thereof or may be obtained from foreign sources. Any leader sequence which is functional in the insect host cell of choice may be used in the present invention.

A polyadenylation sequence may also be operably linked to the 3' terminus of the nucleic acid sequence of the present invention. The polyadenylation sequence is a sequence which when transcribed is recognized by the insect host to add polyadenosine residues to transcribed mRNA. The polyadenylation sequence may be native to the nucleic acid sequence encoding the protein or fragment thereof or may be obtained from foreign sources. Any polyadenylation sequence which is functional in the fungal host of choice may be used in the present invention.

To avoid the necessity of disrupting the cell to obtain the protein or fragment thereof and to minimize the amount of possible degradation of the expressed polypeptide within the cell, it is

preferred that expression of the polypeptide gene gives rise to a product secreted outside the cell. To this end, the protein or fragment thereof of the present invention may be linked to a signal peptide linked to the amino terminus of the protein or fragment thereof. A signal peptide is an amino acid sequence which permits the secretion of the protein or fragment thereof from the insect host into the culture medium. The signal peptide may be native to the protein or fragment thereof of the invention or may be obtained from foreign sources. The 5' end of the coding sequence of the nucleic acid sequence of the present invention may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted protein or fragment thereof.

At present, a mode of achieving secretion of a foreign gene product in insect cells is by way of the foreign gene's native signal peptide. Because the foreign genes are usually from non-insect organisms, their signal sequences may be poorly recognized by insect cells and hence, levels of expression may be suboptimal. However, the efficiency of expression of foreign gene products seems to depend primarily on the characteristics of the foreign protein. On average, nuclear localized or non-structural proteins are most highly expressed, secreted proteins are intermediate and integral membrane proteins are the least expressed. One factor generally affecting the efficiency of the production of foreign gene products in a heterologous host system is the presence of native signal sequences (also termed presequences, targeting signals, or leader sequences) associated with the foreign gene. The signal sequence is generally coded by a DNA sequence immediately following (5' to 3') the translation start site of the desired foreign gene.

The expression dependence on the type of signal sequence associated with a gene product can be represented by the following example: If a foreign gene is inserted at a site downstream from the translational start site of the baculovirus polyhedrin gene so as to produce a fusion

protein (containing the N-terminus of the polyhedrin structural gene), the fused gene is highly expressed. But less expression is achieved when a foreign gene is inserted in a baculovirus expression vector immediately following the transcriptional start site and totally replacing the polyhedrin structural gene.

Insertions into the region -50 to -1 significantly alter (reduce) steady state transcription which, in turn, reduces translation of the foreign gene product. Use of the pVL941 vector optimizes transcription of foreign genes to the level of the polyhedrin gene transcription. Even though the transcription of a foreign gene may be optimal, optimal translation may vary because of several factors involving processing: signal peptide recognition, mRNA and ribosome binding, glycosylation, disulfide bond formation, sugar processing, oligomerization, for example.

The properties of the insect signal peptide are expected to be more optimal for the efficiency of the translation process in insect cells than those from vertebrate proteins. This phenomenon can generally be explained by the fact that proteins secreted from cells are synthesized as precursor molecules containing hydrophobic N-terminal signal peptides. The signal peptides direct transport of the select protein to its target membrane and are then cleaved by a peptidase on the membrane, such as the endoplasmic reticulum, when the protein passes through it.

Another exemplary insect signal sequence is the sequence encoding for *Drosophila* cuticle proteins such as CP1, CP2, CP3 or CP4 (Summers, U.S. Patent No. 5,278,050; the entirety of which is herein incorporated by reference). Most of a 9kb region of the *Drosophila* genome containing genes for the cuticle proteins has been sequenced. Four of the five cuticle genes contains a signal peptide coding sequence interrupted by a short intervening sequence (about 60 base pairs) at a conserved site. Conserved sequences occur in the 5' mRNA

untranslated region, in the adjacent 35 base pairs of upstream flanking sequence and at -200 base pairs from the mRNA start position in each of the cuticle genes.

Standard methods of insect cell culture, cotransfection and preparation of plasmids are set forth in Summers and Smith (Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Agricultural Experiment Station Bulletin No. 1555, Texas A&M University (1987)). Procedures for the cultivation of viruses and cells are described in Volkman and Summers, *J. Virol.* 19:820-832 (1975) and Volkman *et al.*, *J. Virol.* 19:820-832 (1976); both of which are herein incorporated by reference in their entirety.

**(e) Bacterial Constructs and Transformed Bacterial Cells**

The present invention also relates to a bacterial recombinant vector comprising exogenous genetic material. The present invention also relates to a bacteria cell comprising a bacterial recombinant vector. The present invention also relates to methods for obtaining a recombinant bacteria host cell, comprising introducing into a bacterial host cell exogenous genetic material. In a preferred embodiment the exogenous genetic material includes a nucleic acid molecule of the present invention having a sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 677 or complements thereof or fragments of either or other nucleic acid molecule of the present invention.

The bacterial recombinant vector may be any vector which can be conveniently subjected to recombinant DNA procedures. The choice of a vector will typically depend on the compatibility of the vector with the bacterial host cell into which the vector is to be introduced. The vector may be a linear or a closed circular plasmid. The vector system may be a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the bacterial host. In addition, the bacterial vector may be an



expression vector. Nucleic acid molecules encoding protein homologues or fragments thereof can, for example, be suitably inserted into a replicable vector for expression in the bacterium under the control of a suitable promoter for bacteria. Many vectors are available for this purpose and selection of the appropriate vector will depend mainly on the size of the nucleic acid to be inserted into the vector and the particular host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the particular host cell with which it is compatible. The vector components for bacterial transformation generally include, but are not limited to, one or more of the following: a signal sequence, an origin of replication, one or more marker genes and an inducible promoter.

In general, plasmid vectors containing replicon and control sequences that are derived from species compatible with the host cell are used in connection with bacterial hosts. The vector ordinarily carries a replication site, as well as marking sequences that are capable of providing phenotypic selection in transformed cells. For example, *E. coli* is typically transformed using pBR322, a plasmid derived from an *E. coli* species (see, e.g., Bolivar *et al.*, *Gene* 2:95 (1977); the entirety of which is herein incorporated by reference). pBR322 contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed cells. The pBR322 plasmid, or other microbial plasmid or phage, also generally contains, or is modified to contain, promoters that can be used by the microbial organism for expression of the selectable marker genes.

Nucleic acid molecules encoding protein or fragments thereof may be expressed not only directly, but also as a fusion with another polypeptide, preferably a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature polypeptide. In

general, the signal sequence may be a component of the vector, or it may be a part of the polypeptide DNA that is inserted into the vector. The heterologous signal sequence selected should be one that is recognized and processed (i.e., cleaved by a signal peptidase) by the host cell. For bacterial host cells that do not recognize and process the native polypeptide signal sequence, the signal sequence is substituted by a bacterial signal sequence selected, for example, from the group consisting of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Generally, in cloning vectors this sequence is one that enables the vector to replicate independently of the host chromosomal DNA and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria.

Expression and cloning vectors also generally contain a selection gene, also termed a selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*. One example of a selection scheme utilizes a drug to arrest growth of a host cell. Those cells that are successfully transformed with a heterologous protein homologue or fragment thereof produce a protein conferring drug resistance and thus survive the selection regimen.

The expression vector for producing a protein or fragment thereof can also contains an inducible promoter that is recognized by the host bacterial organism and is operably linked to the nucleic acid encoding, for example, the nucleic acid molecule encoding the protein homologue or fragment thereof of interest. Inducible promoters suitable for use with bacterial hosts include the  $\beta$ -lactamase and lactose promoter systems (Chang *et al.*, *Nature* 275:615 (1978); Goeddel *et al.*, *Nature* 281:544 (1979); both of which are herein incorporated by reference in their entirety), the arabinose promoter system (Guzman *et al.*, *J. Bacteriol.* 174:7716-7728 (1992); the entirety of which is herein incorporated by reference), alkaline phosphatase, a tryptophan (trp) promoter system (Goeddel, *Nucleic Acids Res.* 8:4057 (1980); EP 36,776; both of which are herein incorporated by reference in their entirety) and hybrid promoters such as the tac promoter (deBoer *et al.*, *Proc. Natl. Acad. Sci. (USA)* 80:21-25 (1983); the entirety of which is herein incorporated by reference). However, other known bacterial inducible promoters are suitable (Siebenlist *et al.*, *Cell* 20:269 (1980); the entirety of which is herein incorporated by reference).

Promoters for use in bacterial systems also generally contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the polypeptide of interest. The promoter can be removed from the bacterial source DNA by restriction enzyme digestion and inserted into the vector containing the desired DNA.

Construction of suitable vectors containing one or more of the above-listed components employs standard ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored and re-ligated in the form desired to generate the plasmids required. Examples of available bacterial expression vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as Bluescript™ (Stratagene, La Jolla, CA), in which, for example,

encoding an *A. nidulans* protein homologue or fragment thereof homologue, may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke and Schuster, *J. Biol. Chem.* 264:5503-5509 (1989), the entirety of which is herein incorporated by reference); and the like. pGEX vectors (Promega, Madison Wisconsin U.S.A.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems are designed to include heparin, thrombin or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

Suitable host bacteria for a bacterial vector include archaebacteria and eubacteria, especially eubacteria and most preferably *Enterobacteriaceae*. Examples of useful bacteria include *Escherichia*, *Enterobacter*, *Azotobacter*, *Erwinia*, *Bacillus*, *Pseudomonas*, *Klebsiella*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Rhizobia*, *Vitreoscilla* and *Paracoccus*. Suitable *E. coli* hosts include *E. coli* W3110 (American Type Culture Collection (ATCC) 27,325, Manassas, Virginia U.S.A.), *E. coli* 294 (ATCC 31,446), *E. coli* B and *E. coli* X1776 (ATCC 31,537). These examples are illustrative rather than limiting. Mutant cells of any of the above-mentioned bacteria may also be employed. It is, of course, necessary to select the appropriate bacteria taking into consideration replicability of the replicon in the cells of a bacterium. For example, *E. coli*, *Serratia*, or *Salmonella* species can be suitably used as the host when well known plasmids such as pBR322, pBR325, pACYC177, or pKN410 are used to supply the replicon. *E. coli* strain W3110 is a preferred host or parent host because it is a common host strain for recombinant

DNA product fermentations. Preferably, the host cell should secrete minimal amounts of proteolytic enzymes.

Host cells are transfected and preferably transformed with the above-described vectors and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences.

Numerous methods of transfection are known to the ordinarily skilled artisan, for example, calcium phosphate and electroporation. Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in section 1.82 of Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, New York: Cold Spring Harbor Laboratory Press, (1989), is generally used for bacterial cells that contain substantial cell-wall barriers. Another method for transformation employs polyethylene glycol/DMSO, as described in Chung and Miller (Chung and Miller, *Nucleic Acids Res.* 16:3580 (1988); the entirety of which is herein incorporated by reference). Yet another method is the use of the technique termed electroporation.

Bacterial cells used to produce the polypeptide of interest for purposes of this invention are cultured in suitable media in which the promoters for the nucleic acid encoding the heterologous polypeptide can be artificially induced as described generally, e.g., in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, New York: Cold Spring Harbor Laboratory Press, (1989). Examples of suitable media are given in U.S. Pat. Nos. 5,304,472 and 5,342,763; both of which are incorporated by reference in their entirety.

In addition to the above discussed procedures, practitioners are familiar with the standard resource materials which describe specific conditions and procedures for the construction,

manipulation and isolation of macromolecules (e.g., DNA molecules, plasmids, etc.), generation of recombinant organisms and the screening and isolating of clones, (see for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989); Mailga *et al.*, *Methods in Plant Molecular Biology*, Cold Spring Harbor Press (1995), the entirety of which is herein incorporated by reference; Birren *et al.*, *Genome Analysis: Analyzing DNA*, 1, Cold Spring Harbor, New York, the entirety of which is herein incorporated by reference).

**(f) Computer Readable Media**

The nucleotide sequence provided in SEQ ID NO: 1 through SEQ ID NO: 677 or fragment thereof, or complement thereof, or a nucleotide sequence at least 90% identical, preferably 95%, identical even more preferably 99% or 100% identical to the sequence provided in SEQ ID NO: 1 through SEQ ID NO: 677 or fragment thereof, or complement thereof, can be “provided” in a variety of mediums to facilitate use. Such a medium can also provide a subset thereof in a form that allows a skilled artisan to examine the sequences.

A preferred subset of nucleotide sequences are those nucleic acid sequences that encode a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or complement thereof or fragment

of either, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or complement thereof or fragment of either and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or complement thereof or fragment of either.

A further preferred subset of nucleic acid sequences is where the subset of sequences is two proteins or fragments thereof, more preferably three proteins or fragments thereof and even more preferable four transcription factors or fragments thereof, these nucleic acid sequences are selected from the group that comprises a nucleic acid molecule that encodes a putative maize or soybean chlorophyll synthetase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a putative maize or soybean protochlorophyllide reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean coproporphyrinogen oxidase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean protoporphyrinogen oxidase enzyme or complement thereof or fragment of either, a

nucleic acid molecule that encodes a maize or soybean uroporphyrinogen decarboxylase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a putative maize uroporphyrinogen decarboxylase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean porphobilinogen synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean hydroxymethylbilane synthase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate-1-semialdehyde 2,1-aminomutase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamate tRNA ligase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean glutamyl-tRNA reductase enzyme or complement thereof or fragment of either, a nucleic acid molecule that encodes a maize or soybean Mg-chelatase enzyme or complement thereof or fragment of either and a nucleic acid molecule that encodes a maize or soybean ferrochelatase enzyme or complement thereof or fragment of either.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc, storage medium and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention.



As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate media comprising the nucleotide sequence information of the present invention. A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing one or more of nucleotide sequences of the present invention, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990), the entirety of which is herein incorporated by reference) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993), the entirety of which is herein incorporated by reference) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) within the genome that

contain homology to ORFs or proteins from other organisms. Such ORFs are protein-encoding fragments within the sequences of the present invention and are useful in producing commercially important proteins such as enzymes used in amino acid biosynthesis, metabolism, transcription, translation, RNA processing, nucleic acid and a protein degradation, protein modification and DNA replication, restriction, modification, recombination and repair.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify commercially important fragments of the nucleic acid molecule of the present invention. As used herein, "a computer-based system" refers to the hardware means, software means and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As indicated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory that can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention. As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments

or regions of the sequence of the present invention that match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are available can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTIN and BLASTIX (NCBIA). One of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that during searches for commercially important fragments of the nucleic acid molecules of the present invention, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequences the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, *cis* elements, hairpin structures and inducible expression elements (protein binding sequences).

Thus, the present invention further provides an input means for receiving a target sequence, a data storage means for storing the target sequences of the present invention sequence identified using a search means as described above and an output means for outputting the identified homologous sequences. A variety of structural formats for the input and output means

can be used to input and output information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the sequence of the present invention by varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments sequence of the present invention. For example, implementing software which implement the BLAST and BLAZE algorithms (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990)) can be used to identify open frames within the nucleic acid molecules of the present invention. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention.

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration and are not intended to be limiting of the present invention, unless specified.

### **Example 1**

The MONN01 cDNA library is a normalized library generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) total leaf tissue at the V6 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting

at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 6-leaf development stage. The older, more juvenile leaves, which are in a basal position, as well as the younger, more adult leaves, which are more apical are cut at the base of the leaves. The leaves are then pooled and immediately transferred to liquid nitrogen containers in which the pooled leaves are crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON001 cDNA library is generated from maize (B73, Illinois Foundation Seeds, Champaign, Illinois U.S.A.) immature tassels at the V6 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in a greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue from the maize plant is collected at the V6 stage. At that stage the tassel is an immature tassel of about 2-3 cm in length. The tassels are removed and frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON003 library is generated from maize (B73 x Mo17, Illinois Foundation Seeds, Champaign, Illinois U.S.A.) roots at the V6 developmental stage. Seeds are planted at a depth of approximately 3 cm in coil into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth, the seedlings are transplanted into 10 inch pots containing the Metro 200 growing medium. Plants are watered daily before transplantation and approximately 3 times a week after transplantation. Peters 15-16-17 fertilizer is applied approximately three times per week after transplanting at a concentration of 150 ppm N. Two to three times during the life time of the plant from transplanting to flowering a total of approximately 900 mg Fe is added to each pot. Maize plants are grown in the green house in approximately 15hr day/9hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 6 leaf development stage. The root system is cut from maize plant and washed with water to free it from the soil. The tissue is then immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON004 cDNA library is generated from maize (B73 x Mo17, Illinois Foundation Seeds, Champaign, Illinois U.S.A.) total leaf tissue at the V6 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is

approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 6-leaf development stage. The older, more juvenile leaves, which are in a basal position, as well as the younger, more adult leaves, which are more apical are cut at the base of the leaves. The leaves are then pooled and immediately transferred to liquid nitrogen containers in which the pooled leaves are crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON005 cDNA library is generated from maize (B73 x Mo17, Illinois Foundation Seeds, Champaign Illinois, U.S.A.) root tissue at the V6 development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 6-leaf development stage. The root system is cut from the mature maize plant and washed with water to free it from the soil. The tissue is immediately frozen in liquid nitrogen and the harvested tissue is then stored at -80°C until RNA preparation.

The SATMON006 cDNA library is generated from maize (B73 x Mo17, Illinois Foundation Seeds, Champaign Illinois, U.S.A.) total leaf tissue at the V6 plant development

stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 6-leaf development stage. The older more juvenile leaves, which are in a basal position, as well as the younger more adult leaves, which are more apical are cut at the base of the leaves. The leaves are then pooled and immediately transferred to liquid nitrogen containers in which the pooled leaves are crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON007 cDNA library is generated from the primary root tissue of 5 day old maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) seedlings. Seeds are planted on a moist filter paper on a covered tray that is kept in the dark until germination (one day). After germination, the trays, along with the moist paper, are moved to a greenhouse where the maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles for approximately 5 days. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. The primary root tissue is collected when the seedlings are 5 days old. At this stage, the primary root (radicle) is



pushed through the coleorhiza which itself is pushed through the seed coat. The primary root, which is about 2-3 cm long, is cut and immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation.

The SATMON008 cDNA library is generated from the primary shoot (coleoptile 2-3 cm) of maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) seedlings which are approximately 5 days old. Seeds are planted on a moist filter paper on a covered tray that is kept in the dark until germination (one day). Then the trays containing the seeds are moved to a greenhouse at 15hr daytime/9 hr nighttime cycles and grown until they are 5 days post germination. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Tissue is collected when the seedlings are 5 days old. At this stage, the primary shoot (coleoptile) is pushed through the seed coat and is about 2-3 cm long. The coleoptile is dissected away from the rest of the seedling, immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation.

The SATMON009 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) leaves at the 8 leaf stage (V8 plant development stage). Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is 80°F and the nighttime temperature is

70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 8-leaf development stage. The older more juvenile leaves, which are in a basal position, as well as the younger more adult leaves, which are more apical, are cut at the base of the leaves. The leaves are then pooled and then immediately transferred to liquid nitrogen containers in which the pooled leaves are crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON010 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) root tissue at the V8 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is 80°F and the nighttime temperature is 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the V8 development stage. The root system is cut from this mature maize plant and washed with water to free it from the soil. The tissue is immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON011 cDNA library is generated from undeveloped maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) leaf at the V6 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium.

After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 6-leaf development stage. The second youngest leaf which is at the base of the apical leaf of V6 stage maize plant is cut at the base and immediately transferred to liquid nitrogen containers in which the leaf is crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON012 cDNA library is generated from 2 day post germination maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) seedlings. Seeds are planted on a moist filter paper on a covered tray that is kept in the dark until germination (one day). Then the trays containing the seeds are moved to the greenhouse and grown at 15hr daytime/9 hr nighttime cycles until 2 days post germination. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Tissue is collected when the seedlings are 2 days old. At the two day stage, the coleorhiza is pushed through the seed coat and the primary root (the radicle) is pierced the coleorhiza but is barely visible. Also, at this two day stage, the coleoptile is just emerging from the seed coat. The 2 days post germination seedlings are then immersed in liquid nitrogen and crushed. The harvested tissue is stored at -80°C until preparation of total RNA.

660270" 8122260

The SATMON013 cDNA library is generated from apical maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) meristem founder at the V4 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Prior to tissue collection, the plant is at the 4 leaf stage. The lead at the apex of the V4 stage maize plant is referred to as the meristem founder. This apical meristem founder is cut, immediately frozen in liquid nitrogen and crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON014 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) endosperm fourteen days after pollination. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9

hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. After the V10 stage, the maize plant ear shoots are ready for fertilization. At this stage, the ear shoots are enclosed in a paper bag before silk emergence to withhold the pollen. The ear shoots are pollinated and 14 days after pollination, the ears are pulled out and then the kernels are plucked out of the ears. Each kernel is then dissected into the embryo and the endosperm and the aleurone layer is removed. After dissection, the endosperms are immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation.

The SATMON016 library is a maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) sheath library collected at the V8 developmental stage. Seeds are planted in a depth of approximately 3 cm in solid into 2-3 inch pots containing Metro growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and approximately the times a week after transplantation. Peters 15-16-17 fertilizer is applied approximately three times per week after transplanting, at a strength of 150 ppm N. Two to three times during the life time of the plant from transplanting to flowering, a total of approximately 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15hr day/9hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. When the maize plants are at the V8 stage the 5<sup>th</sup> and 6<sup>th</sup> leaves from the bottom exhibit fully developed leaf blades. At the base of these leaves, the ligule is differentiated and the leaf blade is joined to the sheath. The sheath is dissected away from the

base of the leaf then the sheath is frozen in liquid nitrogen and crushed. The tissue is then stored at -80°C until RNA preparation.

The SATMON017 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) embryo seventeen days after pollination. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth the seeds are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. After the V10 stage, the ear shoots of maize plant, which are ready for fertilization, are enclosed in a paper bag before silk emergence to withhold the pollen. The ear shoots are fertilized and 21 days after pollination, the ears are pulled out and the kernels are plucked out of the ears. Each kernel is then dissected into the embryo and the endosperm and the aleurone layer is removed. After dissection, the embryos are immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation.

The SATMON019 (Lib3054) cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) culm (stem) at the V8 developmental stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing

medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. When the maize plant is at the V8 stage, the 5th and 6th leaves from the bottom have fully developed leaf blades. The region between the nodes of the 5th and the sixth leaves from the bottom is the region of the stem that is collected. The leaves are pulled out and the sheath is also torn away from the stem. This stem tissue is completely free of any leaf and sheath tissue. The stem tissue is then frozen in liquid nitrogen and stored at -80°C until RNA preparation.

The SATMON020 cDNA library is from a maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) Hill Type II-Initiated Callus. Petri plates containing approximately 25 ml of Type II initiation media are prepared. This medium contains N6 salts and vitamins, 3% sucrose, 2.3 g/liter proline 0.1 g/liter enzymatic casein hydrolysate, 2mg/liter 2,4 – dichloro phenoxy-acetic acid (2,4, D), 15.3 mg/liter AgNO<sub>3</sub> and 0.8% bacto agar and is adjusted to pH 6.0 before autoclaving. At 9-11 days after pollination, an ear with immature embryos measuring approximately 1-2 mm in length is chosen. The husks and silks are removed and then the ear is broken into halves and placed in an autoclaved solution of Clorox/TWEEN 20 sterilizing solution. Then the ear is rinsed with deionized water. Then each embryo is extracted from the kernel. Intact embryos are placed in contact with the medium, scutellar side up). Multiple embryos are plated on each plate and the plates are incubated in the dark at 25°C. Type II

calluses are friable, can be subcultured with a spatula, frequently regenerate via somatic embryogenesis and are relatively undifferentiated. As seen in the microscope, the Type II calluses show color ranging from translucent to light yellow and heterogeneity on with respect to embryoid structure as well as stage of embryoid development. Once Type II callus are formed, the calluses is transferred to type II callus maintenance medium without  $\text{AgNO}_3$ . Every 7-10 days, the callus is subcultured. About 4 weeks after embryo isolation the callus is removed from the plates and then frozen in liquid nitrogen. The harvested tissue is stored at  $-80^\circ\text{C}$  until RNA preparation.

The SATMON021 cDNA library is generated from the immature maize (DK604, Dekalb Genetics, Dekalb Illinois, U.S.A.) tassel at the V8 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is approximately  $80^\circ\text{F}$  and the nighttime temperature is approximately  $70^\circ\text{F}$ . Supplemental lighting is provided by 1000 W sodium vapor lamps. As the maize plant enters the V8 stage, tassels which are 15-20 cm in length are collected and frozen in liquid nitrogen. The harvested tissue is stored at  $-80^\circ\text{C}$  until RNA preparation.

The SATMON022 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) ear (growing silks) at the V8 plant development stage. Seeds are planted



at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. *Zea mays* plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the plant is in the V8 stage. At this stage, some immature ear shoots are visible. The immature ear shoots (approximately 1 cm in length) are pulled out, frozen in liquid nitrogen and then stored at -80°C until RNA preparation.

The SATMON23 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) ear (growing silk) at the V8 development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. When the tissue is harvested at the V8 stage, the length of the ear that is harvested is about 10-15 cm and the silks are just exposed (approximately 1 inch).

The ear along with the silks is frozen in liquid nitrogen and then the tissue is stored at -80°C until RNA preparation.

The SATMON024 cDNA library is generated from the immature maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) tassel at the V9 development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. As a maize plant enters the V9 stage, the tassel is rapidly developing and a 37 cm tassel along with the glume, anthers and pollen is collected and frozen in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation.

The SATMON025 cDNA library is from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) Hill Type II-Regenerated Callus. Type II callus is grown in initiation media as described for SATMON020 and then the embryoids on the surface of the Type II callus are allowed to mature and germinate. The 1-2 gm fresh weight of the soft friable type callus containing numerous embryoids are transferred to 100 x 15 mm petri plates containing 25 ml of regeneration media. Regeneration media consists of Murashige and Skoog (MS) basal salts, modified White's vitamins (0.2 g/liter glycine and 0.5 g/liter myo-inositol and 0.8% bacto agar (6SMS0D)). The plates are then placed in the dark after covering with parafilm. After 1 week,

the plates are moved to a lighted growth chamber with 16 hr light and 8 hr dark photoperiod. Three weeks after plating the Type II callus to 6SMS0D, the callus exhibit shoot formation. The callus and the shoots are transferred to fresh 6SMS0D plates for another 2 weeks. The callus and the shoots are then transferred to petri plates with reduced sucrose (3SMS0D). Upon distinct formation of a root and shoot, the newly developed green plants are then removed out with a spatula and frozen in liquid nitrogen containers. The harvested tissue is then stored at  $-80^{\circ}\text{C}$  until RNA preparation.

The SATMON026 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) juvenile/adult shift leaves at the V8 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is approximately  $80^{\circ}\text{F}$  and the nighttime temperature is approximately  $70^{\circ}\text{F}$ . Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plants are at the 8-leaf development stage. Leaves are founded sequentially around the meristem over weeks of time and the older, more juvenile leaves arise earlier and in a more basal position than the younger, more adult leaves, which are in a more apical position. In a V8 plant, some leaves which are in the middle portion of the plant exhibit characteristics of both juvenile as well as adult leaves. They exhibit a yellowing color

but also exhibit, in part, a green color. These leaves are termed juvenile/adult shift leaves. The juvenile/adult shift leaves (the 4th, 5th leaves from the bottom) are cut at the base, pooled and transferred to liquid nitrogen in which they are then crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON027 cDNA library is generated from 6 day maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) leaves. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the Metro 200 growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. *Zea mays* plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Prior to tissue collection, when the plant is at the 8-leaf stage, water is held back for six days. The older, more juvenile leaves, which are in a basal position, as well as the younger, more adult leaves, which are more apical, are all cut at the base of the leaves. All the leaves exhibit significant wilting. The leaves are then pooled and immediately transferred to liquid nitrogen containers in which the pooled leaves are then crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON028 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) roots at the V8 developmental stage that are subject to six days water

stress. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the Metro 200 growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Prior to tissue collection, when the plant is at the 8-leaf stage, water is held back for six days. The root system is cut, shaken and washed to remove soil. Root tissue is then pooled and immediately transferred to liquid nitrogen containers in which the pooled leaves are then crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON029 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) seedlings at the etiolated stage. Seeds are planted on a moist filter paper on a covered tray that is kept in the dark for 4 days at approximately 70°F. Tissue is collected when the seedlings are 4 days old. By 4 days, the primary root has penetrated the coleorhiza and is about 4-5 cm and the secondary lateral roots have also made their appearance. The coleoptile has also pushed through the seed coat and is about 4-5 cm long. The seedlings are frozen in liquid nitrogen and crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON030 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) root tissue at the V4 plant development stage. Seeds are planted at a

depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10 inch pots containing the same. Plants are watered daily before transplantation and approximately 3 times a week after transplantation. Peters 15-16-17 fertilizer is applied approximately three times per week after transplanting, at a strength of 150 ppm N. Two to three times during the life time of the plant, from transplanting to flowering, a total of approximately 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15hr day/9hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 sodium vapor lamps. Tissue is collected when the maize plant is at the 4 leaf development stage. The root system is cut from the mature maize plant and washed with water to free it from the soil. The tissue is then immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON031 cDNA library is generated from the maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) leaf tissue at the V4 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is 80°F and the nighttime temperature is 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected

when the maize plant is at the 4-leaf development stage. The third leaf from the bottom is cut at the base and immediately frozen in liquid nitrogen and crushed. The tissue is immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON033 cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) embryo tissue 13 days after pollination. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. After the V10 stage, the ear shoots of the maize plant, which are ready for fertilization, are enclosed in a paper bag before silk emergent to withhold the pollen. The ear shoots are pollinated and 13 days after pollination, the ears are pulled out and then the kernels are plucked cut of the ears. Each kernel is then dissected into the embryo and the endosperm and the aleurone layer is removed. After dissection, the embryos are immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation.

The SATMON034 cDNA library is generated from cold stressed maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) seedlings. Seeds are planted on a moist filter paper on a covered tray that is kept on at 10°C for 7 days. After 7 days, the temperature is shifted to 15°C

for one day until germination of the seed. Tissue is collected once the seedlings are 1 day old. At this point, the coleorhiza has just pushed out of the seed coat and the primary root is just making its appearance. The coleoptile has not yet pushed completely through the seed coat and is also just making its appearance. These 1 day old cold stressed seedlings are frozen in liquid nitrogen and crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMON~001 (Lib36, Lib83, Lib84) cDNA library is generated from maize leaves at the V8 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in a greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue from the maize plant is collected at the V8 stage. The older more juvenile leaves in a basal position as well as the younger more adult leaves which are more apical are all cut at the base, pooled and frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SATMONN01 cDNA library is generated from maize (B73, Illinois Foundation Seeds, Champaign, Illinois U.S.A.) normalized immature tassels at the V6 plant development stage normalized tissue. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into



10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in a greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue from the maize plant is collected at the V6 stage. At that stage the tassel is an immature tassel of about 2-3 cm in length. The tassels are removed and frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation. Single stranded and double stranded DNA representing approximately  $1 \times 10^6$  colony forming units are isolated using standard protocols. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar ratio) and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. The non-hybridized single stranded molecules remaining after hybrid capture are converted to double stranded form and represent the primary normalized library.

The SATMONN04 cDNA library is generated from maize (B73 x Mo17, Illinois Foundation Seeds, Champaign, Illinois U.S.A.) normalized total leaf tissue at the V6 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10

inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 6-leaf development stage. The older, more juvenile leaves, which are in a basal position, as well as the younger, more adult leaves, which are more apical are cut at the base of the leaves. The leaves are then pooled and immediately transferred to liquid nitrogen containers in which the pooled leaves are crushed. The harvested tissue is then stored at -80°C until RNA preparation. Single stranded and double stranded DNA representing approximately  $1 \times 10^6$  colony forming units are isolated using standard protocols. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar ratio) and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. The non-hybridized single stranded molecules remaining after hybrid capture are converted to double stranded form and represent the primary normalized library.

The SATMONN05 cDNA library is generated from maize (B73 x Mo17, Illinois Foundation Seeds, Champaign Illinois, U.S.A.) normalized root tissue at the V6 development

stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 6-leaf development stage. The root system is cut from the mature maize plant and washed with water to free it from the soil. The tissue is immediately frozen in liquid nitrogen and the harvested tissue is then stored at -80°C until RNA preparation. The single stranded and double stranded DNA representing approximately  $1 \times 10^6$  colony forming units are isolated using standard protocols. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar ratio) and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. The non-hybridized single stranded molecules remaining after hybrid capture are converted to double stranded form and represent the primary normalized library.

The SATMONN06 cDNA library is generated from maize (B73 x Mo17, Illinois Foundation Seeds, Champaign Illinois, U.S.A.) normalized total leaf tissue at the V6 plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 6-leaf development stage. The older more juvenile leaves, which are in a basal position, as well as the younger more adult leaves, which are more apical are cut at the base of the leaves. The leaves are then pooled and immediately transferred to liquid nitrogen containers in which the pooled leaves are crushed. The harvested tissue is then stored at -80°C until RNA preparation. Single stranded and double stranded DNA representing approximately  $1 \times 10^6$  colony forming units are isolated using standard protocols. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar ratio) and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. The non-

hybridized single stranded molecules remaining after hybrid capture are converted to double stranded form and represent the primary normalized library.

The CMZ029 (SATMON036) cDNA library is generated from maize (DK604, Dekalb Genetics, Dekalb, Illinois U.S.A.) endosperm 22 days after pollination. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the green house in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. After the V10 stage, the ear shoots of the maize plant, which are ready for fertilization, are enclosed in a paper bag before silk emergent to withhold the pollen. The ear shoots are pollinated and 22 days after pollination, the ears are pulled out and then the kernels are plucked out of the ears. Each kernel is then dissected into the embryo and the endosperm and the alurone layer is removed. After dissection, the endosperms are immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation.

The CMz030 (Lib143) cDNA library is generated from maize seedling tissue two days post germination. Seeds are planted on a moist filter paper on a covered try that is keep in the dark until germination. The trays are then moved to the bench top at 15 hr daytime/9 hr nighttime cycles for 2 days post-germination. The day time temperature is 80°F and the

nighttime temperature is 70°F. Tissue is collected when the seedlings are 2 days old. At this stage, the colehrhiza has pushed through the seed coat and the primary root (the radicle) is just piercing the colehrhiza and is barely visible. The seedlings are placed at 42°C for 1 hour. Following the heat shock treatment, the seedlings are immersed in liquid nitrogen and crushed. The harvested tissue is stored at -80°C until RNA preparation.

The CMz031 (Lib148) cDNA library is generated from maize pollen tissue at the V10+ plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from V10+ stage plants. The ear shoots, which are ready for fertilization, are enclosed in a paper bag to withhold pollen. Twenty-one days after pollination, prior to removing the ears, the paper bag is shaken to collect the mature pollen. The mature pollen is immediately frozen in liquid nitrogen containers and the pollen is crushed. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz033 (Lib189) cDNA library is generated from maize pooled leaf tissue. Samples are harvested from open pollinated plants. Tissue is collected from maize leaves at the

anthesis stage. The leaves are collected from 10-12 plants and frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz034 (Lib3060) cDNA library is generated from maize mature tissue at 40 days post pollination plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from leaves located two leaves below the ear leaf. This sample represents those genes expressed during onset and early stages of leaf senescence. The leaves are pooled and immediately transferred to liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz035 (Lib3061) cDNA library is generated from maize endosperm tissue at the V10+ plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during

the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from V10+ stage plants. The ear shoots, which are ready for fertilization, are enclosed in a paper bag prior to silk emergence to withhold pollen. Thirty-two days after pollination, the ears are pulled out and the kernels are removed from the cob. Each kernel is dissected into the embryo and the endosperm and the aleurone layer is removed. After dissection, the endosperms are immediately transferred to liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz036 (Lib3062) cDNA library is generated from maize husk tissue at the 8 week old plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from 8 week old plants. The husk is separated from the ear and immediately transferred to liquid nitrogen containers. The harvested tissue is then stored at -80°C until RNA preparation.



The CMz037 (Lib3059) cDNA library is generated from maize pooled kernal at 12-15 days after pollination plant development stage. Sample were collected from field grown material. Whole kernal from hand pollinated (control pollination) are harvested as whole ears and immediately frozen on dry ice. Kernels from 10-12 ears were pooled and ground together in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz039 (Lib3066) cDNA library is generated from maize immature anther tissue at the 7 week old immature tassel stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 7 week old immature tassel stage. At this stage, prior to anthesis, the immature anthers are green and enclosed in the staminate spikelet. The developing anthers are dissected away from the 7 week old immature tassel and immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz040 (Lib3067) cDNA library is generated from maize kernel tissue at the V10+ plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into

10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from V10+ stage plants. The ear shoots, which are ready for fertilization, are enclosed in a paper bag before silk emergence to withhold pollen. Five to eight days after controlled pollination. The ears are pulled and the kernels removed. The kernels are immediately frozen in liquid nitrogen. The harvested kernels tissue is then stored at -80°C until RNA preparation. This sample represents gene expressed in early kernel development, during periods of cell division, amyloplast biogenesis and early carbon flow across the material to filial tissue.

The CMz041 (Lib3068) cDNA library is generated from maize pollen germinating silk tissue at the V10+ plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F.

Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from V10+ stage plants when the ear shoots are ready for fertilization at the silk emergence stage. The emerging silks are pollinated with an excess of pollen under controlled pollination conditions in the green house. Eighteen hours after pollination the silks are removed from the ears and immediately frozen in liquid nitrogen containers. This sample represents genes expressed in both pollen and silk tissue early in pollination. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz042 (Lib3069) cDNA library is generated from maize ear tissue excessively pollinated at the V10+ plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F.

Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from V10+ stage plants and the ear shoots which are ready for fertilization are at the silk emergence stage. The immature ears are pollinated with an excess of pollen under controlled pollination conditions. Eighteen hours post-pollination, the ears are removed and immediately transferred to liquid nitrogen containers. The harvested tissue is then stored at -80°C until RNA preparation.

660270" 8T2E260

The CMz044 (Lib3075) cDNA library is generated from maize microspore tissue at the V10+ plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from immature anthers from 7 week old tassels. The immature anthers are first dissected from the 7 week old tassel with a scalpel on a glass slide covered with water. The microspores (immature pollen) are released into the water and are recovered by centrifugation. The microspore suspension is immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz045 (Lib3076) cDNA library is generated from maize immature ear megaspore tissue. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are

grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from immature ear (megaspore) obtained from 7 week old plants. The immature ears are harvested from the 7 week old plants and are approximately 2.5 to 3 cm in length. The kernels are removed from the cob immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz047 (Lib3078) cDNA library is generated from maize CO<sub>2</sub> treated high-exposure shoot tissue at the V10+ plant development stage. RX601 maize seeds are sterilized for 1 minute with a 10% clorox solution. The seeds are rolled in germination paper, and germinated in 0.5 mM calcium sulfate solution for two days at 30°C. The seedlings are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium at a rate of 2-3 seedlings per pot. Twenty pots are placed into a high CO<sub>2</sub> environment (approximately 1000 ppm CO<sub>2</sub>). Twenty plants were grown under ambient greenhouse CO<sub>2</sub> (approximately 450 ppm CO<sub>2</sub>). Plants are watered daily before transplantation and three times a week after transplantation. Peters 20-20-20 fertilizer is also lightly applied. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. At ten days post planting, the shoots from both atmosphere are frozen in liquid nitrogen and lightly ground. The roots are washed in deionized water to remove the support media and the tissue is immediately transferred to liquid nitrogen containers. The harvested tissue is then stored at -80°C until RNA preparation.

09233218-012099

The CMz048 (Lib3079) cDNA library is generated from maize basal endosperm transfer layer tissue at the V10+ plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected from V10+ maize plants. The ear shoots, which are ready for fertilization, are enclosed in a paper bag prior to silk emergence, to withhold the pollen. Kernels are harvested at 12 days post-pollination and placed on wet ice for dissection. The kernels are cross sectioned laterally, dissecting just above the pedicel region, including 1-2 mm of the lower endosperm and the basal endosperm transfer region. The pedicel and lower endosperm region containing the basal endosperm transfer layer is pooled and immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz049(Lib3088) cDNA library is generated from maize immature anther tissue at the 7 week old immature tassel stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is

applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F.

Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is at the 7 week old immature tassel stage. At this stage, prior to anthesis, the immature anthers are green and enclosed in the staminate spikelet. The developing anthers are dissected away from the 7 week old immature tassel and immediately transferred to liquid nitrogen container. The harvested tissue is then stored at -80°C until RNA preparation.

The CMz050 (Lib3114) cDNA library is generated from maize silk tissue at the V10+ plant development stage. Seeds are planted at a depth of approximately 3 cm into 2-3 inch peat pots containing Metro 200 growing medium. After 2-3 weeks growth they are transplanted into 10 inch pots containing the same growing medium. Plants are watered daily before transplantation and three times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Maize plants are grown in the greenhouse in 15 hr day/9 hr night cycles. The daytime temperature is approximately 80°F and the nighttime temperature is approximately 70°F. Supplemental lighting is provided by 1000 W sodium vapor lamps. Tissue is collected when the maize plant is beyond the 10-leaf development stage and the ear shoots are approximately 15-20 cm in length. The ears are pulled and silks are separated from the ears and immediately

transferred to liquid nitrogen containers. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON001 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) total leaf tissue at the V4 plant development stage. Leaf tissue from 38, field grown V4 stage plants is harvested from the 4<sup>th</sup> node. Leaf tissue is removed from the plants and immediately frozen in dry-ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON002 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) root tissue at the V4 plant development stage. Root tissue from 76, field grown V4 stage plants is harvested. The root systems is cut from the soybean plant and washed with water to free it from the soil and immediately frozen in dry-ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON003 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seedling hypocotyl axis tissue harvested 2 day post-imbibition. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium. Trays are placed in an environmental chamber and grown at 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Tissue is collected 2 days after the start of imbibition. The 2 days after imbibition samples are separated into 3 collections after removal of any adhering seed coat. At the 2 day stage, the hypocotyl axis is emerging from the soil. A few seedlings have cracked the soil surface and exhibited slight greening of the exposed cotyledons. The seedlings are washed



in water to remove soil, hypocotyl axis harvested and immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON004 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seedling cotyledon tissue harvested 2 day post-imbibition. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium. Trays are placed in an environmental chamber and grown at 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Tissue is collected 2 days after the start of imbibition. The 2 days after imbibition samples are separated into 3 collections after removal of any adhering seed coat. At the 2 day stage, the hypocotyl axis is emerging from the soil. A few seedlings have cracked the soil surface and exhibited slight greening of the exposed cotyledons. The seedlings are washed in water to remove soil, hypocotyl axis harvested and immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON005 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seedling hypocotyl axis tissue harvested 6 hour post-imbibition. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium. Trays are placed in an environmental chamber and grown at 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Tissue is collected 6 hours after the start of imbibition. The 6 hours after imbibition samples are separated into 3 collections after removal of any adhering seed coat. The 6 hours after imbibition sample is collected over the course of approximately 2 hours starting at 6

hours post imbibition. At the 6 hours after imbibition stage, not all cotyledons have become fully hydrated and germination, or radicle protrusion, has not occurred. The seedlings are washed in water to remove soil, hypocotyl axis harvested and immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON006 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seedling cotyledons tissue harvest 6 hour post-imbibition. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium. Trays are placed in an environmental chamber and grown at 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Tissue is collected 6 hours after imbibition. The 6 hours after imbibition samples are separated into 3 collections after removal of any adhering seed coat. The 6 hours after imbibition sample is collected over the course of approximately 2 hours starting at 6 hours post-imbibition. At the 6 hours after imbibition, not all cotyledons have become fully hydrated and germination or radicle protrusion, have not occurred. The seedlings are washed in water to remove soil, cotyledon harvested and immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON007 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seed tissue harvested 25 and 35 days post-flowering. Seed pods from field grown plants are harvested 25 and 35 days after flowering and the seeds extracted from the pods. Approximately 4.4g and 19.3g of seeds are harvested from the respective seed pods and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON008 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) leaf tissue harvested from 25 and 35 days post-flowering plants. Total leaf tissue is harvested from field grown plants. Approximately 19g and 29g of leaves are harvested from the fourth node of the plant 25 and 35 days post-flowering and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON009 cDNA library is generated from soybean cutlivar C1944 (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) pod and seed tissue harvested 15 days post-flowering. Pods from field grown plants are harvested 15 days post-flowering. Approximately 3g of pod tissue is harvested and immediately frozen in dry-ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON010 cDNA library is generated from soybean cultivar C1944 (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) seed tissue harvested 40 days post-flowering. Pods from field grown plants are harvested 40 days post-flowering. Pods and seeds are separated, approximately 19g of seed tissue is harvested and immediately frozen in dry-ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON011 cDNA library is generated from soybean cultivars Cristalina (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) and FT108 (Monsoy, Brazil) (tropical germ plasma) leaf tissue. Leaves are harvested from plants grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Approximately 30g of leaves are harvested from the 4<sup>th</sup> node of each

of the Cristalina and FT108 cultivars and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON012 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) leaf tissue. Leaves from field grown plants are harvested from the fourth node 15 days post-flowering. Approximately 12g of leaves are harvested and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON013 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) root and nodule tissue. Approximately, 28g of root tissue from field grown plants is harvested 15 days post-flowering. The root system is cut from the soybean plant, washed with water to free it from the soil and immediately frozen in dry-ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON014 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seed tissue harvested 25 and 35 days after flowering. Seed pods from field grown plants are harvested 15 days after flowering and the seeds extracted from the pods. Approximately 5g of seeds are harvested from the respective seed pods and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON015 cDNA is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seed tissue harvested 45 and 55 days post-flowering. Seed pods from field grown plants are harvested 45 and 55 days after flowering and the seeds extracted from the pods. Approximately 19g and 31g of seeds are harvested from the respective

seed pods and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON016 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) root tissue. Approximately, 61g and 38g of root tissue from field grown plants is harvested 25 and 35 days post- flowering is harvested. The root system is cut from the soybean plant and washed with water to free it from the soil. The tissue is placed in 14ml polystyrene tubes and immediately frozen in dry-ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON017 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) root tissue. Approximately 28g of root tissue from field grown plants is harvested 45 and 55 days post-flowering. The root system is cut from the soybean plant, washed with water to free it from the soil and immediately frozen in dry-ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON018 cDNA is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) leaf tissue harvested 45 and 55 days post-flowering. Leaves from field grown plants are harvested 45 and 55 days after flowering from the fourth node. Approximately 27g and 33g of seeds are harvested from the respective seed pods and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON019 cDNA library is generated from soybean cultivars Cristalina (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) and FT108 (Monsoy, Brazil) (tropical germ plasma) root tissue. Roots are harvested from plants grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and

the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Approximately 50g and 56g of roots are harvested from each of the Cristalina and FT108 cultivars and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON020 cDNA is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seed tissue harvested 65 and 75 days post-flowering. Seed pods from field grown plants are harvested 45 and 55 days after flowering and the seeds extracted from the pods. Approximately 14g and 31g of seeds are harvested from the respective seed pods and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON021 cDNA library is generated from Soybean Cyst Nematode-resistant soybean cultivar Hartwig (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) root tissue. Plants are grown in tissue culture at room temperature. At approximately 6 weeks post-germination, the plants are exposed to sterilized Soybean Cyst Nematode eggs. Infection is then allowed to progress for 10 days. After the 10 day infection process, the tissue is harvested. Agar from the culture medium and nematodes are removed and the root tissue is immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON022 (Lib3030) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) partially opened flower tissue. Partially to fully opened flower tissue is harvested from plants grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to

maintain even moisture conditions. A total of 3g of flower tissue is harvested and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON023 cDNA library is generated from soybean genotype BW211S Null (Tohoku University, Morioka, Japan) seed tissue harvested 15 and 40 days post-flowering. Seed pods from field grown plants are harvested 15 and 40 days post-flowering and the seeds extracted from the pods. Approximately 0.7g and 14.2g of seeds are harvested from the respective seed pods and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON024 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) internode-2 tissue harvested 18 days post-imbibition. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium. The plants are grown in a greenhouse for 18 days after the start of imbibition at ambient temperature. Soil is checked and watered daily to maintain even moisture conditions. Stem tissue is harvested 18 days after the start of imbibition. The samples are divided into hypocotyl and internodes 1 through 5. The fifth internode contains some leaf bud material. Approximately 3g of each sample is harvested and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON025 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) leaf tissue harvested 65 days post-flowering. Leaves are harvested from the fourth node of field grown plants 65 days post-flowering. Approximately 18.4g of leaf tissue is harvested and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

SOYMON026 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) root tissue harvested 65 and 75 days post-flowering. Approximately 27g and 40g of root tissue from field grown plants is harvested 65 and 75 days post- flowering. The root system is cut from the soybean plant, washed with water to free it from the soil and immediately frozen in dry-ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON027 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seed tissue harvested 25 days post-flowering. Seed pods from field grown plants are harvested 25 days post-flowering and the seeds extracted from the pods. Approximately 17g of seeds are harvested from the seed pods and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON028 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) drought-stressed root tissue. The plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature 24°C. Soil is checked and watered daily to maintain even moisture conditions. At the R3 stage of development, water is withheld from half of the plant collection (drought stressed population). After 3 days, half of the plants from the drought stressed condition and half of the plants from the control population are harvested. After another 3 days (6 days post drought induction) the remaining plants are harvested. A total of 27g and 40g of root tissue is harvested and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.



The SOYMON029 cDNA library is generated from Soybean Cyst Nematode-resistant soybean cultivar PI07354 (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) root tissue. Late fall to early winter greenhouse grown plants are exposed to Soybean Cyst Nematode eggs. At 10 days post-infection, the plants are uprooted, rinsed briefly and the roots frozen in liquid nitrogen. Approximately 20 grams of root tissue is harvested from the infected plants. The harvested tissue is then stored at  $-80^{\circ}\text{C}$  until RNA preparation.

The SOYMON030 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) flower bud tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately  $29^{\circ}\text{C}$  and the nighttime temperature approximately  $24^{\circ}\text{C}$ . Soil is checked and watered daily to maintain even moisture conditions. Flower buds are removed from the plant at the pedicel. A total of 100mg of flower buds are harvested and immediately frozen in liquid nitrogen. The harvested tissue is then stored at  $-80^{\circ}\text{C}$  until RNA preparation.

The SOYMON031 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) carpel and stamen tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately  $29^{\circ}\text{C}$  and the nighttime temperature approximately  $24^{\circ}\text{C}$ . Soil is checked and watered daily to maintain even moisture conditions. Flower buds are removed from the plant at the pedicel. Flowers are dissected to separate petals, sepals and reproductive structures (carpels and stamens). A total of 300mg of carpel and stamen tissue are

harvested and immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON032 cDNA library is prepared from the Asgrow cultivar A4922 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) rehydrated dry soybean seed meristem tissue. Surface sterilized seeds are germinated in liquid media for 24 hours. The seed axis is then excised from the barely germinating seed, placed on tissue culture media and incubated overnight at 20°C in the dark. The supportive tissue is removed from the explant prior to harvest. Approximately 570mg of tissue is harvested and frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON033 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) heat-shocked seedling tissue without cotyledons. Seeds are imbibed and germinated in vermiculite for 2 days under constant illumination. After 48 hours, the seedlings are transferred to an incubator set at 40°C under constant illumination. After 30, 60 and 180 minutes seedlings are harvested and dissected. A portion of the seedling consisting of the root, hypocotyl and apical hook is frozen in liquid nitrogen and stored at -80°C. The seedlings after 2 days of imbibition are beginning to emerge from the vermiculite surface. The apical hooks are dark green in appearance. Total RNA and poly A<sup>+</sup> RNA is prepared from equal amounts of pooled tissue.

The SOYMON034 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) cold-shocked seedling tissue without cotyledons. Seeds are imbibed and germinated in vermiculite for 2 days under constant illumination. After 48 hours, the seedlings are transferred to a cold room set at 5°C under constant illumination. After 30, 60 and 180 minutes seedlings are harvested and dissected. A

portion of the seedling consisting of the root, hypocotyl and apical hook is frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . The seedlings after 2 days of imbibition are beginning to emerge from the vermiculite surface. The apical hooks are dark green in appearance.

The SOYMON035 cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seed coat tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately  $29^{\circ}\text{C}$  and the nighttime temperature  $24^{\circ}\text{C}$ . Soil is checked and watered daily to maintain even moisture conditions. Seeds are harvested from mid to nearly full maturation (seed coats are not yellowing). The entire embryo proper is removed from the seed coat sample and the seed coat tissue are harvested and immediately frozen in liquid nitrogen. The harvested tissue is then stored at  $-80^{\circ}\text{C}$  until RNA preparation.

The SOYMON036 cDNA library is generated from soybean cultivars PI171451, PI227687 and PI229358 (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) insect challenged leaves. Plants from each of the three cultivars are grown in screenhouse conditions. The screenhouse is divided in half and one half of the screenhouse is infested with soybean looper and the other half infested with velvetbean caterpillar. A single leaf is taken from each of the representative plants at 3 different time points, 11 days after infestation, 2 weeks after infestation and 5 weeks after infestation and immediately frozen in liquid nitrogen. The harvested tissue is then stored at  $-80^{\circ}\text{C}$  until RNA preparation. Total RNA and poly A<sup>+</sup> RNA is isolated from pooled tissue consisting of equal quantities of all 18 samples (3 genotypes X 3 sample times X 2 insect genotypes).

The SOYMON037 cDNA library is generated from soybean cultivar A3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) etiolated axis and radical tissue. Seeds are planted in moist vermiculite, wrapped and kept at room temperature in complete darkness until harvest. Etiolated axis and hypocotyl tissue is harvested at 2, 3 and 4 days post-planting. A total of 1 gram of each tissue type is harvested at 2, 3 and 4 days after planting and immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The SOYMON038 cDNA library is generated from soybean variety Asgrow A3237 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) rehydrated dry seeds. Explants are prepared for transformation after germination of surface-sterilized seeds on solid tissue media. After 6 days, at 28°C and 18 hours of light per day, the germinated seeds are cold shocked at 4°C for 24 hours. Meristemic tissue and part of the hypocotyl is removed and cotyledon excised. The prepared explant is then wounded for *Agrobacterium* infection. The 2 grams of harvested tissue is frozen in liquid nitrogen and stored at -80°C until RNA preparation.

The Soy51 (LIB3027) cDNA library is prepared from equal amounts tissue harvested from SOYMON007, SOYMON015 and SOYMON020 prepared tissue. Single stranded and double stranded DNA representing approximately  $1 \times 10^6$  colony forming units are isolated using standard protocols. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar ratio) and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. The non-hybridized single

stranded molecules remaining after hybrid capture are converted to double stranded form and represent the primary normalized library.

The Soy52 (LIB3028) cDNA library is generated from normalized flower DNA. Single stranded DNA representing approximately  $1 \times 10^6$  colony forming units of SOYMON022 harvested tissue is used as the starting material for normalization. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar ratio) and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. The non-hybridized single stranded molecules remaining after hybrid capture are converted to double stranded form and represent the primary normalized library.

The Soy53 (LIB3039) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) seedling shoot apical meristem tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature 24°C. Soil is checked and watered daily to maintain even moisture conditions. Apical tissue is harvested from seedling shoot meristem tissue, 7-8 days after the start of imbibition. The apex of each seedling is dissected to include the fifth node to the apical meristem. The fifth node corresponds to the third trifoliate leaf in the very early stages of development. Stipules completely envelop the leaf primordia at this time. A total of 200mg of apical tissue is harvested

and immediately frozen in liquid nitrogen. The harvested tissue is then stored at -80°C until RNA preparation.

The Soy54 (LIB3040) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) heart to torpedo stage embryo tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature 24°C. Soil is checked and watered daily to maintain even moisture conditions. Seeds are collected and embryos removed from surrounding endosperm and maternal tissues. Embryos from globular to young torpedo stages (by corresponding analogy to *Arabidopsis*) are collected with a bias towards the middle of this spectrum. Embryos which are beginning to show asymmetric development of cotyledons are considered the upper developmental boundary for the collection and are excluded. A total of 12 mg embryo tissue is frozen in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation.

Soy55 (LIB3049) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) young seed tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature 24°C. Soil is checked and watered daily to maintain even moisture conditions. Seeds are collected from very young pods (5 to 15 days after flowering). A total of 100mg of seeds are harvested and frozen in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation.

Soy56 (LIB3029) cDNA library is prepared from equal amounts tissue harvested from SOYMON007, SOYMON015 and SOYMON020 prepared tissue. Single stranded and double stranded DNA representing approximately  $1 \times 10^6$  colony forming units are isolated using standard protocols. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar ratio and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. The non-hybridized single stranded molecules remaining after hybrid capture are not converted to double stranded form and represent a non-normalized seed pool for comparison to Soy51 cDNA libraries.

TheSoy58 (LIB3050) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) drought stressed root tissue subtracted from control root tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature 24°C. Soil is checked and watered daily to maintain even moisture conditions. At the R3 stage of the plant drought is induced by withholding water. After 3 and 6 days root tissue from both drought stressed and control (watered regularly) plants are collected and frozen in dry-ice. The harvested tissue is stored at -80°C until RNA preparation. For subtraction, target cDNA is made from the drought stressed tissue total RNA using the SMART cDNA synthesis system from Clontech (Clontech Laboratories, Palo Alto, California U.S.A.). Driver first strand cDNA is covalently linked to Dynabeads following a protocol similar to that

described in the Dynal literature (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The target cDNA is then heat denatured and the second strand trapped using Dynabeads oligo-dT. The target second strand cDNA is then hybridized to the driver cDNA in 400µl 2X SSPE for two rounds of hybridization at 65°C and 20 hours. After each hybridization, the hybridization solution is removed from the system and the hybridized target cDNA removed from the driver by heat denaturation in water. After hybridization, the remaining cDNA is trapped with Dynabeads oligo-dT. The trapped cDNA is then amplified as in previous PCR based libraries and the resulting cDNA ligated into the pSPORT vector (Invitrogen, Carlsbad California U.S.A.).

The Soy59 (LIB3051) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) endosperm tissue. Seeds are germinated on paper towels under laboratory ambient light conditions. At 8, 10 and 14 hours after imbibition, the seed coats are harvested. The endosperm consists of a very thin layer of tissue affixed to the inside of the seed coat. The seed coat and endosperm are frozen immediately after harvest in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation.

The Soy60 (LIB3072) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) drought stressed seed plus pod subtracted from control seed plus pod tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 26°C and the nighttime temperature 21°C and 70% relative humidity. Soil is checked and watered daily to maintain even moisture conditions. At the R3 stage of the plant drought is induced by withholding water. After 3 and 6 days seeds and pods from both drought stressed and



control (watered regularly) plants are collected from the fifth and sixth node and frozen in dry-ice. The harvested tissue is stored at -80°C until RNA preparation. For subtraction, target cDNA is made from the drought stressed tissue total RNA using the SMART cDNA synthesis system from Clontech (Clontech Laboratories, Palo Alto, California U.S.A.). Driver first strand cDNA is covalently linked to Dynabeads following a protocol similar to that described in the Dynal literature (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The target cDNA is then heat denatured and the second strand trapped using Dynabeads oligo-dT. The target second strand cDNA is then hybridized to the driver cDNA in 400 µl 2X SSPE for two rounds of hybridization at 65°C and 20 hours. After each hybridization, the hybridization solution is removed from the system and the hybridized target cDNA removed from the driver by heat denaturation in water. After hybridization, the remaining cDNA is trapped with Dynabeads oligo-dT. The trapped cDNA is then amplified as in previous PCR based libraries and the resulting cDNA ligated into the pSPORT vector (Invitrogen, Carlsbad California U.S.A.).

The Soy61 (LIB3073) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) jasmonic acid treated seedling subtracted from control tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in a greenhouse. The daytime temperature is approximately 29.4°C and the nighttime temperature 20°C. Soil is checked and watered daily to maintain even moisture conditions. At 9 days post planting, the plantlets are sprayed with either control buffer of 0.1% Tween-20 or jasmonic acid (Sigma J-2500, Sigma, St. Louis, Missouri U.S.A.) at 1 mg/ml in 0.1% Tween-20. Plants are sprayed until runoff and the soil and the stem is soaked with the spraying solution. At 18 hours post application of jasmonic acid, the soybean plantlets appear growth retarded. After 18hours, 24hours and 48 hours post

treatment, the cotyledons are removed and the remaining leaf and stem tissue above the soil is harvested and frozen in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation. To make RNA, the three sample timepoints were combined and ground. For subtraction, target cDNA is made from the jasmonic acid treated tissue total RNA using the SMART cDNA synthesis system from Clontech (Clontech Laboratories, Palo Alto, California U.S.A.). Driver first strand cDNA is covalently linked to Dynabeads following a protocol similar to that described in the Dynal literature (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The target cDNA is then heat denatured and the second strand trapped using Dynabeads oligo-dT. The target second strand cDNA is then hybridized to the driver cDNA in 400 µl 2X SSPE for two rounds of hybridization at 65°C and 20 hours. After each hybridization, the hybridization solution is removed from the system and the hybridized target cDNA removed from the driver by heat denaturation in water. After hybridization, the remaining cDNA is trapped with Dynabeads oligo-dT. The trapped cDNA is then amplified as in previous PCR based libraries and the resulting cDNA ligated into the pSPORT vector (Invitrogen, Carlsbad California U.S.A.). For this library's construction, the eighth fraction of the cDNA size fractionation step was used for ligation.

The Soy62 (LIB3074) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) jasmonic acid treated seedling subtracted from control tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in a greenhouse. The daytime temperature is approximately 29.4°C and the nighttime temperature 20°C. Soil is checked and watered daily to maintain even moisture conditions. At 9 days post planting, the plantlets are sprayed with either control buffer of 0.1% Tween-20 or jasmonic acid (Sigma J-2500, Sigma, St.

Loius, Missouri U.S.A.) at 1 mg/ml in 0.1% Tween-20. Plants are sprayed until runoff and the soil and the stem is soaked with the spraying solution. At 18 hours post application of jasmonic acid, the soybean plantlets appear growth retarded. After 18 hours, 24 hours and 48 hours post treatment, the cotyledons are removed and the remaining leaf and stem tissue above the soil is harvested and frozen in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation. To make RNA, the three sample timepoints were combined and ground. For subtraction, target cDNA is made from the jasmonic acid treated tissue total RNA using the SMART cDNA synthesis system from Clontech (Clontech Laboratories, Palo Alto, California U.S.A.). Driver first strand cDNA is covalently linked to Dynabeads following a protocol similar to that described in the Dynal literature (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The target cDNA is then heat denatured and the second strand trapped using Dynabeads oligo-dT. The target second strand cDNA is then hybridized to the driver cDNA in 400 µl 2X SSPE for two rounds of hybridization at 65°C and 20 hours. After each hybridization, the hybridization solution is removed from the system and the hybridized target cDNA removed from the driver by heat denaturation in water. After hybridization, the remaining cDNA is trapped with Dynabeads oligo-dT. The trapped cDNA is then amplified as in previous PCR based libraries and the resulting cDNA ligated into the pSPORT vector (Invitrogen, Carlsbad California U.S.A.). For this library's construction, the ninth fraction of the cDNA size fractionation step was used for ligation.

The Soy65 (LIB3107) 07cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) drought-stressed abscission zone tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr

nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature 24°C. Soil is checked and watered daily to maintain even moisture conditions. Plants are irrigated with 15-16-17 Peter's Mix. At the R3 stage of development, drought is imposed by withholding water. At 3, 4, 5 and 6 days, tissue is harvested and wilting is not obvious until the fourth day. Abscission layers from reproductive organs are harvested by cutting less than one millimeter proximal and distal to the layer and immediately frozen in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation.

The Soy66 (LIB3109) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) non-drought stressed abscission zone tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Plants are irrigated with 15-16-17 Peter's Mix. At 3, 4, 5 and 6 days, control abscission layer tissue is harvested. Abscission layers from reproductive organs are harvested by cutting less than one millimeter proximal and distal to the layer and immediately frozen in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation.

Soy67 (LIB3065) cDNA library is prepared from equal amounts tissue harvested from SOYMON007, SOYMON015 and SOYMON020 prepared tissue. Single stranded and double stranded DNA representing approximately  $1 \times 10^6$  colony forming units are isolated using standard protocols. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar

ratio) and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. Captured hybrids are eluted with water.

Soy68 (LIB3052) cDNA library is prepared from equal amounts tissue harvested from SOYMON007, SOYMON015 and SOYMON020 prepared tissue. Single stranded and double stranded DNA representing approximately  $1 \times 10^6$  colony forming units are isolated using standard protocols. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar ratio) and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. Captured hybrids are eluted with water.

Soy69 (LIB3053) cDNA library is generated from soybean cultivars Cristalina (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) and FT108 (Monsoy, Brazil) (tropical germ plasma) normalized leaf tissue. Leaves are harvested from plants grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Approximately 30g of leaves are harvested from the 4<sup>th</sup> node of each of the Cristalina and FT108 cultivars and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation. Single stranded and double stranded DNA representing approximately  $1 \times 10^6$  colony forming units are isolated using standard protocols. RNA, complementary to the single stranded DNA, is synthesized using the double stranded DNA as a template. Biotinylated dATP is incorporated into the RNA during the

synthesis reaction. The single stranded DNA is mixed with the biotinylated RNA in a 1:10 molar ratio) and allowed to hybridize. DNA-RNA hybrids are captured on Dynabeads M280 streptavidin (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The dynabeads with captured hybrids are collected with a magnet. The non-hybridized single stranded molecules remaining after hybrid capture are converted to double stranded form and represent the primary normalized library.

Soy70 (LIB3055) cDNA library is generated from soybean cultivars Cristalina (USDA Soybean Germplasm Collection, Urbana, Illinois U.S.A.) and FT108 (Monsoy, Brazil) (tropical germ plasma) leaf tissue. Leaves are harvested from plants grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Approximately 30g of leaves are harvested from the 4<sup>th</sup> node of each of the Cristalina and FT108 cultivars and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

Soy71 (LIB3056) cDNA library is generated from soybean cultivars Cristalina and FT108 (tropical germ plasma) root tissue. Roots are harvested from plants grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 29°C and the nighttime temperature approximately 24°C. Soil is checked and watered daily to maintain even moisture conditions. Approximately 50g and 56g of roots are harvested from each of the Cristalina and FT108 cultivars and immediately frozen in dry ice. The harvested tissue is then stored at -80°C until RNA preparation.

Soy72 (LIB3093) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) drought stressed leaf control tissue. Seeds

are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under 12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 26°C and the nighttime temperature 21°C and 70% relative humidity. Soil is checked and watered daily to maintain even moisture conditions. At the R3 stage of the plant drought is induced by withholding water. After 3 and 6 days seeds and pods from both drought stressed and control (watered regularly) plants are collected from the fifth and sixth node and frozen in dry-ice. The harvested tissue is stored at -80°C until RNA preparation. For subtraction, target cDNA is made from the drought stressed tissue total RNA using the SMART cDNA synthesis system from Clontech (Clontech Laboratories, Palo Alto, California U.S.A.). Driver first strand cDNA is covalently linked to Dynabeads following a protocol similar to that described in the Dynal literature (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The target cDNA is then heat denatured and the second strand trapped using Dynabeads oligo-dT. The target second strand cDNA is then hybridized to the driver cDNA in 400 µl 2X SSPE for two rounds of hybridization at 65°C and 20 hours. After each hybridization, the hybridization solution is removed from the system and the hybridized target cDNA removed from the driver by heat denaturation in water. After hybridization, the remaining cDNA is trapped with Dynabeads oligo-dT. The trapped cDNA is then amplified as in previous PCR based libraries and the resulting cDNA ligated into the pSPORT vector (Invitrogen, Carlsbad California U.S.A.).

Soy73 (LIB3093) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) drought stressed leaf subtracted from control tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in an environmental chamber under

12hr daytime/12hr nighttime cycles. The daytime temperature is approximately 26°C and the nighttime temperature 21°C and 70% relative humidity. Soil is checked and watered daily to maintain even moisture conditions. At the R3 stage of the plant drought is induced by withholding water. After 3 and 6 days seeds and pods from both drought stressed and control (watered regularly) plants are collected from the fifth and sixth node and frozen in dry-ice. The harvested tissue is stored at -80°C until RNA preparation. For subtraction, target cDNA is made from the drought stressed tissue total RNA using the SMART cDNA synthesis system from Clontech (Clontech Laboratories, Palo Alto, California U.S.A.). Driver first strand cDNA is covalently linked to Dynabeads following a protocol similar to that described in the Dynal literature (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The target cDNA is then heat denatured and the second strand trapped using Dynabeads oligo-dT. The target second strand cDNA is then hybridized to the driver cDNA in 400 µl 2X SSPE for two rounds of hybridization at 65°C and 20 hours. After each hybridization, the hybridization solution is removed from the system and the hybridized target cDNA removed from the driver by heat denaturation in water. After hybridization, the remaining cDNA is trapped with Dynabeads oligo-dT. The trapped cDNA is then amplified as in previous PCR based libraries and the resulting cDNA ligated into the pSPORT vector (Invitrogen, Carlsbad California U.S.A.).

The Soy76 (Lib3106) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) jasmonic acid and arachidonic treated seedling subtracted from control tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in a greenhouse. The daytime temperature is approximately 29.4°C and the nighttime temperature 20°C. Soil is checked and watered daily to maintain even moisture conditions. At 9 days post planting, the



plantlets are sprayed with either control buffer of 0.1% Tween-20 or jasmonic acid (Sigma J-2500, Sigma, St. Louis, Missouri U.S.A.) at 1 mg/ml in 0.1% Tween-20. Plants are sprayed until runoff and the soil and the stem is soaked with the spraying solution. At 18 hours post application of jasmonic acid, the soybean plantlets appear growth retarded. Arachidonic treated seedlings are sprayed with 1m/ml arachidonic acid in 0.1% Tween-20. After 18hours, 24hours and 48 hours post treatment, the cotyledons are removed and the remaining leaf and stem tissue above the soil is harvested and frozen in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation. To make RNA, the three sample timepoints were combined and ground. The RNA from the arachidonic treated seedlings is isolated separately. For subtraction, target cDNA is made from the jasmonic acid treated tissue total RNA using the SMART cDNA synthesis system from Clontech (Clontech Laboratories, Palo Alto, California U.S.A.). Driver first strand cDNA is covalently linked to Dynabeads following a protocol similar to that described in the Dynal literature (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The target cDNA is then heat denatured and the second strand trapped using Dynabeads oligo-dT. The target second strand cDNA is then hybridized to the driver cDNA in 400 µl 2X SSPE for two rounds of hybridization at 65°C and 20 hours. After each hybridization, the hybridization solution is removed from the system and the hybridized target cDNA removed from the driver by heat denaturation in water. After hybridization, the remaining cDNA is trapped with Dynabeads oligo-dT. The trapped cDNA is then amplified as in previous PCR based libraries and the resulting cDNA ligated into the pSPORT vector (Invitrogen, Carlsbad California U.S.A.). Fraction 10 of the size fractionated cDNA is ligated into the pSPORT vector (Invitrogen, Carlsbad California U.S.A.) in order to capture some of the smaller transcripts characteristic of antifungal proteins.

Soy77 (LIB3108) cDNA library is generated from soybean cultivar Asgrow 3244 (Asgrow Seed Company, Des Moines, Iowa U.S.A.) jasmonic acid control tissue. Seeds are planted at a depth of approximately 2cm into 2-3 inch peat pots containing Metromix 350 medium and the plants are grown in a greenhouse. The daytime temperature is approximately 29.4°C and the nighttime temperature 20°C. Soil is checked and watered daily to maintain even moisture conditions. At 9 days post planting, the plantlets are sprayed with either control buffer of 0.1% Tween-20 or jasmonic acid (Sigma J-2500, Sigma, St. Louis, Missouri U.S.A.) at 1 mg/ml in 0.1% Tween-20. Plants are sprayed until runoff and the soil and the stem is soaked with the spraying solution. At 18 hours post application of jasmonic acid, the soybean plantlets appear growth retarded. Arachidonic treated seedlings are sprayed with 1m/ml arachidonic acid in 0.1% Tween-20. After 18 hours, 24 hours and 48 hours post treatment, the cotyledons are removed and the remaining leaf and stem tissue above the soil is harvested and frozen in liquid nitrogen. The harvested tissue is stored at -80°C until RNA preparation. To make RNA, the three sample timepoints were combined and ground. The RNA from the arachidonic treated seedlings is isolated separately. For subtraction, target cDNA is made from the jasmonic acid treated tissue total RNA using the SMART cDNA synthesis system from Clontech (Clontech Laboratories, Palo Alto, California U.S.A.). Driver first strand cDNA is covalently linked to Dynabeads following a protocol similar to that described in the Dynal literature (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.). The target cDNA is then heat denatured and the second strand trapped using Dynabeads oligo-dT. The target second strand cDNA is then hybridized to the driver cDNA in 400 µl 2X SSPE for two rounds of hybridization at 65°C and 20 hours. After each hybridization, the hybridization solution is removed from the system and the hybridized target cDNA removed from the driver by heat denaturation in water. After

hybridization, the remaining cDNA is trapped with Dynabeads oligo-dT. The trapped cDNA is then amplified as in previous PCR based libraries and the resulting cDNA ligated into the pSPORT vector (Invitrogen, Carlsbad California U.S.A.). Fraction 10 of the size fractionated cDNA is ligated into the pSPORT vector in order to capture some of the smaller transcripts characteristic of antifungal proteins.

The Lib9 cDNA library is prepared from *Arabidopsis thaliana*, Columbia ecotype, leaf tissue. Wild type *Arabidopsis thaliana* seeds are planted in commonly used planting pots and grown in an environmental chamber. Leaf blades were cut with sharp scissors at seven weeks after planting. The tissue was immediately frozen in liquid nitrogen. The harvested tissue is stored at  $-80^{\circ}\text{C}$  until RNA extraction. PolyA mRNA is purified from the total RNA preparation using Dynabeads® Oligo(dT)<sub>25</sub> (DynaL Inc., Lake Success, N.Y.), or equivalent methods. This library was normalized using a PCR-based protocol.

The Lib22 cDNA library is prepared from *Arabidopsis thaliana* Columbia ecotype, root tissue. Wild type *Arabidopsis thaliana* seeds are planted in commonly used planting pots and grown in an environmental chamber. After 5-6 weeks the plants are in the reproductive growth phase. Stems are bolting from the base of the plants. After 7 weeks, more stems, floral buds appear, and a few flowers are starting to open. The 7-week old plants are rinsed intensively by tope water remove dirt from the roots, and blotted by paper towel. The tissues are immediately frozen in liquid nitrogen. The harvested tissue is stored at  $-80^{\circ}\text{C}$  until RNA preparation.

The Lib23 cDNA library is prepared from *Arabidopsis thaliana*, Columbia ecotype, stem tissue. Wild type *Arabidopsis thaliana* seeds are planted in commonly used planting pots and grown in an environmental chamber. Stems were collected seven to eight weeks after planting

by cutting the stems from the base and cutting the top of the plant to remove the floral tissue. The tissue was immediately frozen in liquid nitrogen and stored at -80°C until total RNA extraction. PolyA mRNA is purified from the total RNA preparation using Dynabeads® Oligo(dT)<sub>25</sub> (DynaI Inc., Lake Success, N.Y.), or equivalent methods. This library was normalized using a PCR-based protocol.

The Lib24 cDNA library is prepared from *Arabidopsis thaliana*, Columbia ecotype, flower bud tissue. Wild type *Arabidopsis thaliana* seeds are planted in commonly used planting pots and grown in an environmental chamber. Flower buds are green and unopened and harvested about seven weeks after planting. The tissue is immediately frozen in liquid nitrogen. The harvested tissue is stored at -80°C until total RNA extraction. PolyA mRNA is purified from the total RNA preparation using Dynabeads® Oligo(dT)<sub>25</sub> (DynaI Inc., Lake Success, N.Y.), or equivalent methods. This library was normalized using a PCR-based protocol.

The Lib25 cDNA library is prepared from *Arabidopsis thaliana*, Columbia ecotype, open flower tissue. Wild type *Arabidopsis thaliana* seeds are planted in commonly used planting pots and grown in an environmental chamber. Flowers are completely opened with all parts of floral structure observable, but no siliques are appearing. The tissue was immediately frozen in liquid nitrogen and stored at -80°C until total RNA extraction. PolyA mRNA is purified from the total RNA preparation using Dynabeads® Oligo(dT)<sub>25</sub> (DynaI Inc., Lake Success, N.Y.), or equivalent methods. This library was normalized using a PCR-based protocol.

The Lib35 cDNA library of the present invention, was prepared from *Arabidopsis thaliana* Columbia ecotype leaf tissue. Wild type *Arabidopsis thaliana* seeds are planted in commonly used planting pots and grown in an environmental chamber. After 5-6 weeks the

plants are in the reproductive growth phase. Stems are bolting from the base of the plants. After 7 weeks, more stems and floral buds appeared and a few flowers were starting to open. Leaf blades were collected by cutting with sharp scissors. The tissues were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until use. PolyA mRNA is purified from the total RNA preparation using Dynabeads® Oligo(dT)<sub>25</sub> (DynaL Inc., Lake Success, N.Y.), or equivalent methods. This library was normalized using a PCR-based protocol.

The Lib146 cDNA library is prepared from *Arabidopsis thaliana*, Columbia ecotype, immature seed tissue. Wild type *Arabidopsis thaliana* seeds are planted in commonly used planting pots and grown in an environmental chamber. At approximately 7-8 weeks of age, the seeds are harvested. The seeds ranged in maturity from the smallest seeds that could be dissected from silques to just before starting to turn yellow in color. The tissue is immediately frozen in liquid nitrogen. The harvested tissue is stored at  $-80^{\circ}\text{C}$  until RNA extraction. PolyA mRNA is purified from the total RNA preparation using Dynabeads® Oligo(dT)<sub>25</sub> (DynaL Inc., Lake Success, N.Y.), or equivalent methods. This library is normalized using a PCR-based protocol.

The Lib3032 (Lib80) cDNA libraries are generated from *Brassica napus* seeds harvested 30 days after pollination. The cDNA libraries are constructed using the SuperScript Plasmid system for cDNA synthesis and plasmid cloning (Life Technologies, Gaithersburg, Maryland U.S.A.) according to the manufacturers protocol with the following modification: 40 micrograms of total RNA is used as the starting material for cDNA synthesis, and first strand cDNA synthesis is carried out at  $45^{\circ}\text{C}$ .

The Lib3034 (Lib82) cDNA libraries are generated from *Brassica napus* seeds harvested 15 and 18 days after pollination. The cDNA libraries are constructed using the SuperScript

Plasmid system for cDNA synthesis and plasmid cloning (Life Technologies, Gaithersburg, Maryland U.S.A.) according to the manufacturers protocol with the following modification: 40 micrograms of total RNA is used as the starting material for cDNA synthesis, and first strand cDNA synthesis was carried out at 45°C.

The Lib3099 cDNA library is generated by a subtraction procedure. The library contains cDNAs whose abundance is enriched in the *Brassica napus* 15 and 18 day after pollination seed tissues when compared to *Brassica* leaf tissues. The cDNA synthesis is performed on *Brassica* leaf RNA and *Brassica* RNA isolated from seeds harvested 15 and 18 days after pollination using a Smart PCR cDNA synthesis kit according to the manufacturers protocol (Clontech, Palo Alto, California U.S.A.). The subtracted cDNA is generated using the Clontech PCR-Select subtraction kit according to the manufacturers protocol (Clontech, Palo Alto, California U.S.A.). The subtracted cDNA was cloned into plasmid vector pCR2.1 according to the manufacturers protocol (Invitrogen, Carlsbad, California U.S.A.).

The Lib3033 (Lib81) cDNA libraries are generated from from the *Schizochytrium* species cells. The *Schizochytrium* species cells are grown in liquid media until saturation. The culture is centrifuged to pellet the cells, the medium is decanted off, and pellet immediately frozen in liquid nitrogen. Wax esters are produced under such dark, anaerobic, rich-medium conditions. High wax production by the cultures is verified by microscopy (fluorescein staining of wax bodies) and by lipid extraction/TLC/GC. The harvested cells are stored at -80°C until RNA preparation. RNA is prepared from the frozen *Euglena* cell pellet as follows. The pellet is pulverized to a powder in liquid nitrogen using a mortar and pestle. The powder is transferred to tubes containing 6 ml each of lysis buffer (100 mM Tris, pH 8, 0.6 M NaCl, 10 mM EDTA, and 4% (w/v) SDS) and buffered phenol, vortexed, and disrupted with a Polytron. The mixture is

centrifuged 20 min at 10,000xg in Corex glass tubes to separate the phases. 5 ml of the upper phase is removed, vortexed with 5 ml fresh phenol, and centrifuged. The upper phase is removed and the RNA is precipitated overnight at 4°C by adding 1.5 volumes of 4 M LiCl. The RNA is further purified on Rneasy columns according to the manufacturers protocol (Qiagen, Valencia, California U.S.A.). The cDNA library is constructed using the SuperScript Plasmid system for cDNA synthesis and plasmid cloning (Life Technologies, Gaithersburg, Maryland U.S.A.) according to the manufacturers protocol with the following modification: 40 micrograms of total RNA was used as the starting material for cDNA synthesis, and first strand cDNA synthesis was carried out at 45°C.

The Lib47 cDNA library is generated from *Euglena gracilis* strain 753 (ATTC No. 30285, ATCC Manasas, Virginia U.S.A.) grown in liquid culture. A liquid culture is inoculated with 1/10 volume of a previously-grown saturated culture, and the new culture for 4 days under near-anaerobic conditions (near-anaerobic cultures are not agitated, just gently swirled once a day) in the dark in 2X Beef (10 g/l bacto peptone, 4 g/l yeast extract, 2 g/l beef extract, 6 g/l glucose). The culture is then centrifuged to pellet the cells, the medium is decanted off, and pellet immediately frozen in liquid nitrogen. Wax esters are produced under such dark, anaerobic, rich-medium conditions. High wax production by the cultures is verified by microscopy (fluorescein staining of wax bodies) and by lipid extraction/TLC/GC. The harvested cells are stored at -80°C until RNA preparation. RNA is prepared from the frozen *Euglena* cell pellet as follows. The pellet is pulverized to a powder in liquid nitrogen using a mortar and pestle. The powder is transferred to tubes containing 6 ml each of lysis buffer (100 mM Tris, pH 8, 0.6 M NaCl, 10 mM EDTA, and 4% (w/v) SDS) and buffered phenol, vortexed, and disrupted with a Polytron. The mixture is centrifuged 20 min at 10,000xg in Corex glass tubes to separate

the phases. 5 ml of the upper phase is removed, vortexed with 5 ml fresh phenol, and centrifuged. The upper phase is removed and the RNA is precipitated overnight at 4°C by adding 1.5 volumes of 4 M LiCl. The RNA is further purified on Rneasy columns according to the manufacturers protocol (Qiagen, Valencia, California U.S.A.). The cDNA library is constructed using the SuperScript Plasmid system for cDNA synthesis and plasmid cloning (Life Technologies, Gaithersburg, Maryland U.S.A.) according to the manufacturers protocol with the following modification: 40 micrograms of total RNA was used as the starting material for cDNA synthesis, and first strand cDNA synthesis was carried out at 45°C.

The Lib44 cDNA library is generated from *Phaeodactylum tricornutum* grown in modified Jones medium for 3 days. The cells were harvested by centrifugation and the resulting pellet frozen immediately in liquid nitrogen. The harvested cells are stored at -80°C until RNA preparation. RNA is prepared from the frozen *Phaeodactylum* cell pellet as follows. The pellet is pulverized to a powder in liquid nitrogen using a mortar and pestle. The powder is transferred to tubes containing 6 ml each of lysis buffer (100 mM Tris, pH 8, 0.6 M NaCl, 10 mM EDTA, and 4% (w/v) SDS) and buffered phenol, vortexed, and disrupted with a Polytron. The mixture is centrifuged 20 min at 10,000xg in Corex glass tubes to separate the phases. 5 ml of the upper phase is removed, vortexed with 5 ml fresh phenol, and centrifuged. The upper phase is removed and the RNA is precipitated overnight at 4°C by adding 1.5 volumes of 4 M LiCl. The RNA is further purified on Rneasy columns according to the manufacturers protocol (Qiagen, Valencia, California U.S.A.). The cDNA library is constructed using the SuperScript Plasmid system for cDNA synthesis and plasmid cloning (Life Technologies, Gaithersburg, Maryland U.S.A.) according to the manufacturers protocol with the following modification: 40 micrograms of total



RNA was used as the starting material for cDNA synthesis, and first strand cDNA synthesis was carried out at 45 degrees centigrade.

The LIB3036 genomic library is generated from *Mycobacterium neoaurum* US52 (ATCC No. 23072, ATCC, Manassas, Virginia U.S.A.) cells. *Mycobacterium neoaurum* US52 is a gram-positive Actinomycete bacterium. *Mycobacterium neoaurum* US52 is genetically related to *Mycobacterium tuberculosis*, but there is no reason to believe that it is a primary pathogen. It normally is saprophytic, i.e. it lives in soil and gets nutrients from decaying matter. Genomic DNA obtained from *Mycobacterium neoaurum* US52 is digested for various times with the restriction enzyme Sau3A. The DNA fractions are size-separated on an agarose gel, and the first fraction wherein most of the partially-digested fragments are about 10 kB is used to isolated fragments in the range of 2-3 kB. For LIB3036, the 2-3 kB DNA is cloned into vector pRY401 (Invitrogen, Carlsbad, California U.S.A.). The vector pZERO-2 (Invitrogen, Carlsbad, California U.S.A.) is used for the construction of LIB3104.

The stored RNA is purified using Trizol reagent from Life Technologies (Gibco BRL, Life Technologies, Gaithersburg, Maryland U.S.A.), essentially as recommended by the manufacturer. Poly A+ RNA (mRNA) is purified using magnetic oligo dT beads essentially as recommended by the manufacturer (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.).

Construction of plant cDNA libraries is well-known in the art and a number of cloning strategies exist. A number of cDNA library construction kits are commercially available. The Superscript™ Plasmid System for cDNA synthesis and Plasmid Cloning (Gibco BRL, Life

Technologies, Gaithersburg, Maryland U.S.A.) is used, following the conditions suggested by the manufacturer.

Normalized libraries are made using essentially the Soares procedure (Soares *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 91:9228-9232 (1994), the entirety of which is herein incorporated by reference). This approach is designed to reduce the initial 10,000-fold variation in individual cDNA frequencies to achieve abundances within one order of magnitude while maintaining the overall sequence complexity of the library. In the normalization process, the prevalence of high-abundance cDNA clones decreases dramatically, clones with mid-level abundance are relatively unaffected and clones for rare transcripts are effectively increased in abundance.

### **Example 2**

The cDNA libraries are plated on LB agar containing the appropriate antibiotics for selection and incubated at 37° for a sufficient time to allow the growth of individual colonies. Single colonies are individually placed in each well of a 96-well microtiter plates containing LB liquid including the selective antibiotics. The plates are incubated overnight at approximately 37°C with gentle shaking to promote growth of the cultures. The plasmid DNA is isolated from each clone using Qiaprep plasmid isolation kits, using the conditions recommended by the manufacturer (Qiagen Inc., Santa Clara, California U.S.A.).

Template plasmid DNA clones are used for subsequent sequencing. For sequencing, the ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq® DNA Polymerase, FS, is used (PE Applied Biosystems, Foster City, California U.S.A.).

### **Example 3**

Nucleic acid sequences that encode for the following tetrapyrrole pathway enzymes: putative chlorophyll synthetase, protochlorophyllide reductase, putative protochlorophyllide reductase, coproporphyrinogen oxidase, protoporphyrinogen oxidase, uroporphyrinogen decarboxylase, putative uroporphyrinogen decarboxylase, porphobilinogen synthase enzyme, hydroxymethylbilane synthase enzyme, glutamate-1-semialdehyde 2,1-aminomutase enzyme, glutamate tRNA ligase enzyme, glutamyl-tRNA reductase enzyme, Mg-chelatase enzyme, and ferrochelatase enzyme are identified from the Monsanto EST PhytoSeq database using TBLASTN (default values)(TBLASTN compares a protein query against the six reading frames of a nucleic acid sequence). Matches found with BLAST P values equal or less than 0.001 (probability) or BLAST Score of equal or greater than 90 are classified as hits. If the program used to determine the hit is HMMSW then the score refers to HMMSW score.

In addition, the GenBank database is searched with BLASTN and BLASTX (default values) using ESTs as queries. EST that pass the hit probability threshold of  $10e^{-8}$  for the following enzymes are combined with the hits generated by using TBLASTN (described above) and classified by enzyme (see Table A below).

A cluster refers to a set of overlapping clones in the PhytoSeq database. Such an overlapping relationship among clones is designated as a “cluster” when BLAST scores from pairwise sequence comparisons of the member clones meets a predetermined minimum value or product score of 50 or more (Product Score = (BLAST SCORE x Percentage Identity)/(5 x minimum [length (Seq1), length (Seq2)]))

Since clusters are formed on the basis of single-linkage relationships, it is possible for two non-overlapping clones to be members of the same cluster if, for instance, they both overlap

a third clone with at least the predetermined minimum BLAST score (stringency). A cluster ID is arbitrarily assigned to all of those clones which belong to the same cluster at a given stringency and a particular clone will belong to only one cluster at a given stringency. If a cluster contains only a single clone (a “singleton”), then the cluster ID number will be negative, with an absolute value equal to the clone ID number of its single member. Clones grouped in a cluster in most cases represent a contiguous sequence.

**TABLE A\***

**SOYBEAN PUTATIVE CHLOROPHYLL SYNTHETASE**

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
1	-700941050	700941050H1	SOYMON024	g972938	BLASTX	349	1e-41	75
2	-701212263	701212263H1	SOYMON035	g972938	BLASTX	75	1e-9	65
3	-701213734	701213734H1	SOYMON035	g972937	BLASTN	191	1e-27	83
4	14458	LIB3049-005-Q1-E1-F12	LIB3049	g3068709	BLASTX	101	1e-35	59
5	14458	700975706H1	SOYMON009	g972938	BLASTX	75	1e-9	50
6	14458	701047496H1	SOYMON032	g972938	BLASTX	75	1e-9	52
7	26375	701156709H1	SOYMON031	g972938	BLASTX	102	1e-15	92
8	26375	701156060H1	SOYMON031	g972937	BLASTN	275	1e-13	80

**SOYBEAN PROTOCHLOROPHYLLIDE REDUCTASE**

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
9	-700654876	700654876H1	SOYMON004	g20829	BLASTN	269	1e-23	82
10	-700657235	700657235H1	SOYMON004	g20829	BLASTN	728	1e-57	83
11	-700657437	700657437H1	SOYMON004	g20829	BLASTN	668	1e-46	84
12	-700757662	700757662H1	SOYMON015	g20829	BLASTN	1012	1e-83	89
13	-700842232	700842232H1	SOYMON020	g20829	BLASTN	442	1e-32	81
14	-700976426	700976426H1	SOYMON009	g2244613	BLASTN	1038	1e-77	85
15	11407	700652980H1	SOYMON003	g2244613	BLASTN	741	1e-52	71
16	11407	700735503H1	SOYMON010	g2244613	BLASTN	505	1e-33	70
17	11407	701142847H1	SOYMON038	g2244613	BLASTN	509	1e-33	67
18	11407	700652452H1	SOYMON003	g2244613	BLASTN	535	1e-33	69
19	11407	700735307H1	SOYMON010	g2244613	BLASTN	491	1e-32	70
20	11407	701107638H1	SOYMON036	g2244613	BLASTN	493	1e-32	70
21	11407	700952613H1	SOYMON022	g2244613	BLASTN	479	1e-31	70
22	11407	701118520H1	SOYMON037	g2244613	BLASTN	475	1e-29	70
23	11407	700731513H1	SOYMON010	g2244613	BLASTN	455	1e-28	72
24	11407	701037153H1	SOYMON029	g2244613	BLASTN	455	1e-27	72
25	11407	700838406H1	SOYMON020	g2244614	BLASTX	229	1e-24	57
26	11407	700736971H1	SOYMON010	g2244613	BLASTN	389	1e-22	75
27	11407	701208151H1	SOYMON035	g2244613	BLASTN	387	1e-21	76
28	11407	700658204H1	SOYMON004	g2244614	BLASTX	173	1e-16	61
29	11407	700657759H1	SOYMON004	g2244614	BLASTX	120	1e-14	54
30	11407	700854307H1	SOYMON023	g20829	BLASTN	190	1e-12	80
31	2160	LIB3039-002-Q1-E1-G10	LIB3039	g2244613	BLASTN	690	1e-46	84
32	2160	701107175H1	SOYMON036	g20829	BLASTN	628	1e-43	88
33	21731	700660488H1	SOYMON004	g20829	BLASTN	796	1e-59	86
34	21731	701134573H1	SOYMON038	g20829	BLASTN	646	1e-44	84
35	21739	700655688H1	SOYMON004	g2244613	BLASTN	298	1e-35	84
36	21739	700655588H1	SOYMON004	g20830	BLASTX	125	1e-21	88
37	2977	700763883H1	SOYMON018	g20829	BLASTN	892	1e-80	83
38	2977	701139350H1	SOYMON038	g20829	BLASTN	614	1e-72	88
39	2977	700849172H1	SOYMON021	g20829	BLASTN	956	1e-70	87
40	2977	700993334H1	SOYMON011	g20829	BLASTN	793	1e-69	87
41	2977	700980689H1	SOYMON009	g20829	BLASTN	765	1e-66	83
42	2977	700754834H1	SOYMON014	g20829	BLASTN	910	1e-66	86
43	2977	701054679H1	SOYMON032	g20829	BLASTN	501	1e-60	81

44	2977	701142549H1	SOYMON038	g20829	BLASTN	595	1e-59	82
45	2977	700909828H1	SOYMON022	g20829	BLASTN	695	1e-59	84
46	2977	701153047H1	SOYMON031	g20829	BLASTN	715	1e-50	87
47	2977	700981305H1	SOYMON009	g2244613	BLASTN	645	1e-44	70
48	2977	700737910H1	SOYMON012	g2244613	BLASTN	598	1e-41	69
49	2977	701106762H1	SOYMON036	g2244613	BLASTN	602	1e-41	70
50	2977	700893019H1	SOYMON024	g2244613	BLASTN	595	1e-40	70
51	2977	700888819H1	SOYMON024	g2244613	BLASTN	555	1e-37	69
52	2977	700557617H1	SOYMON001	g2244613	BLASTN	557	1e-37	69
53	2977	700989268H1	SOYMON011	g20829	BLASTN	297	1e-35	80
54	2977	700978858H1	SOYMON009	g2244613	BLASTN	529	1e-35	69
55	2977	701063251H1	SOYMON033	g2244613	BLASTN	525	1e-34	63
56	2977	700737989H1	SOYMON012	g20829	BLASTN	190	1e-33	72
57	2977	LIB3054-001- Q1-B1-A11	LIB3054	g2244613	BLASTN	487	1e-32	70
58	2977	701057704H1	SOYMON033	g2244613	BLASTN	470	1e-29	72
59	2977	701139740H1	SOYMON038	g2244613	BLASTN	477	1e-29	69
60	2977	LIB3039-043- Q1-E1-F3	LIB3039	g2244614	BLASTX	99	1e-28	55
61	2977	701105971H1	SOYMON036	g2244613	BLASTN	454	1e-28	71
62	2977	700789775H1	SOYMON011	g2244613	BLASTN	429	1e-26	72
63	2977	700732675H1	SOYMON010	g2244613	BLASTN	437	1e-26	69
64	2977	701137164H1	SOYMON038	g2244613	BLASTN	429	1e-25	72
65	2977	700788180H1	SOYMON011	g2244613	BLASTN	431	1e-25	72
66	2977	700680942H1	SOYMON008	g20829	BLASTN	349	1e-24	72
67	2977	700953017H1	SOYMON022	g2244613	BLASTN	395	1e-22	71
68	2977	700962368H1	SOYMON022	g2244613	BLASTN	395	1e-22	71
69	2977	700737258H1	SOYMON010	g2244613	BLASTN	395	1e-22	71
70	2977	701058308H1	SOYMON033	g2244613	BLASTN	244	1e-14	78
71	2977	701108820H1	SOYMON036	g968974	BLASTN	254	1e-14	76
72	2977	700658246H1	SOYMON004	g20830	BLASTX	123	1e-13	76
73	2977	700990646H1	SOYMON011	g20829	BLASTN	255	1e-12	92
74	2977	700548092H1	SOYMON001	g20830	BLASTX	92	1e-11	79
75	2977	701136902H1	SOYMON038	g2244613	BLASTN	265	1e-11	69
76	2977	701152877H1	SOYMON031	g20830	BLASTX	128	1e-10	76
77	2977	700994862H1	SOYMON011	g20830	BLASTX	128	1e-10	76
78	2977	701148824H1	SOYMON031	g20830	BLASTX	128	1e-10	76
79	2977	701047440H1	SOYMON032	g20830	BLASTX	128	1e-10	76
80	2977	700556683H1	SOYMON001	g968974	BLASTN	252	1e-10	76
81	2977	701146931H1	SOYMON031	g2244613	BLASTN	253	1e-10	72
82	2977	701142178H1	SOYMON038	g20830	BLASTX	122	1e-9	73
83	2977	701152593H1	SOYMON031	g20830	BLASTX	123	1e-9	75
84	2977	700737725H1	SOYMON012	g20829	BLASTN	218	1e-9	74
85	2977	700683007H1	SOYMON008	g2244613	BLASTN	241	1e-9	70
86	2977	700739072H1	SOYMON012	g2244613	BLASTN	244	1e-9	71
87	4903	700658027H1	SOYMON004	g20829	BLASTN	820	1e-59	79
88	4903	700852934H1	SOYMON023	g20829	BLASTN	453	1e-48	78
89	6970	LIB3052-012- Q1-N1-A11	LIB3052	g968974	BLASTN	934	1e-69	78
90	6970	700660679H1	SOYMON004	g20829	BLASTN	862	1e-67	87
91	6970	700682420H2	SOYMON008	g968976	BLASTN	864	1e-63	80
92	6970	700979758H2	SOYMON009	g2244613	BLASTN	865	1e-63	82
93	6970	700790842H1	SOYMON011	g968974	BLASTN	642	1e-57	82
94	6970	700994812H1	SOYMON011	g968976	BLASTN	423	1e-43	78

### SOYBEAN PUTATIVE PROTOCHLOROPHYLLIDE REDUCTASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
95	-701065431	701065431H1	SOYMON034	g348719	BLASTN	767	1e-55	83
96	4640	700982771H1	SOYMON009	g348718	BLASTX	162	1e-15	93

### SOYBEAN COPROPORHYRINOGEN OXIDASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
97	-700671956	700671956H1	SOYMON006	g414665	BLASTN	291	1e-16	96
98	-701053612	701053612H1	SOYMON032	g414665	BLASTN	335	1e-27	94
99	-701208513	701208513H1	SOYMON035	g414665	BLASTN	639	1e-92	94
100	11665	700656318H1	SOYMON004	g414665	BLASTN	656	1e-93	98
101	11665	700964466H1	SOYMON022	g414665	BLASTN	611	1e-88	98
102	11665	700899782H1	SOYMON027	g414665	BLASTN	648	1e-87	98
103	11665	700844365H1	SOYMON021	g414665	BLASTN	648	1e-83	98
104	11665	701146220H1	SOYMON031	g414665	BLASTN	630	1e-75	98
105	11665	700660179H1	SOYMON004	g414665	BLASTN	530	1e-55	94
106	11665	700662658H1	SOYMON005	g1213066	BLASTN	742	1e-53	78
107	11665	701152413H1	SOYMON031	g1213066	BLASTN	742	1e-53	78
108	6121	LIB3065-002-Q1-N1-G8	LIB3065	g414665	BLASTN	1383	1e-128	96
109	6121	701108945H1	SOYMON036	g414665	BLASTN	1388	1e-106	98
110	6121	700789601H2	SOYMON011	g414665	BLASTN	1301	1e-99	99
111	6121	700994436H1	SOYMON011	g414665	BLASTN	858	1e-97	98
112	6121	700747416H1	SOYMON013	g414665	BLASTN	925	1e-94	100
113	6121	700978804H1	SOYMON009	g414665	BLASTN	941	1e-89	95
114	6121	701109318H1	SOYMON036	g414665	BLASTN	729	1e-85	96
115	6121	700873742H1	SOYMON018	g414665	BLASTN	484	1e-82	94
116	6121	701209226H1	SOYMON035	g414665	BLASTN	789	1e-81	97
117	6121	701060931H1	SOYMON033	g414665	BLASTN	768	1e-79	96
118	6121	701066887H1	SOYMON034	g414665	BLASTN	262	1e-68	90
119	6121	700899224H1	SOYMON027	g414665	BLASTN	434	1e-57	84
120	6121	700906273H1	SOYMON022	g414665	BLASTN	555	1e-37	97
121	6121	700992008H1	SOYMON011	g414665	BLASTN	560	1e-37	97
122	6121	700786848H2	SOYMON011	g414665	BLASTN	408	1e-25	98
123	6121	700734585H1	SOYMON010	g414665	BLASTN	250	1e-16	100
124	7272	700786276H2	SOYMON011	g414665	BLASTN	1170	1e-88	97
125	7272	700683451H1	SOYMON008	g414665	BLASTN	1019	1e-87	97
126	7272	700662424H1	SOYMON005	g414665	BLASTN	932	1e-84	98
127	7882	700680869H1	SOYMON008	g414665	BLASTN	763	1e-54	99
128	7882	700680628H1	SOYMON008	g414665	BLASTN	516	1e-34	94

### SOYBEAN PROTOPORPHYRINOGEN OXIDASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
129	-700657957	700657957H1	SOYMON004	g1183006	BLASTN	729	1e-51	76
130	-700681258	700681258H1	SOYMON008	g1183006	BLASTN	651	1e-46	76
131	-701063830	701063830H1	SOYMON034	g2370335	BLASTX	142	1e-15	79

### SOYBEAN UROPORPHYRINOGEN DECARBOXYLASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
132	-700730557	700730557H1	SOYMON009	g1009428	BLASTN	444	1e-38	69
133	-700789740	700789740H1	SOYMON011	g1009428	BLASTN	760	1e-54	80
134	-700974704	700974704H1	SOYMON005	g1016347	BLASTX	272	1e-30	55
135	-701048641	701048641H1	SOYMON032	g1009427	BLASTN	580	1e-39	71
136	-GM17920	LIB3055-003-Q1-N1-H10	LIB3055	g142136	BLASTX	97	1e-29	61
137	19517	701104233H1	SOYMON036	g1009429	BLASTX	228	1e-24	49
138	19517	701000103H1	SOYMON018	g1009429	BLASTX	167	1e-22	46
139	19517	701108875H1	SOYMON036	g1009429	BLASTX	137	1e-19	49
140	19517	700737952H1	SOYMON012	g1009429	BLASTX	188	1e-18	39
141	4729	700753974H1	SOYMON014	g1009427	BLASTN	816	1e-59	82
142	4729	701126044H1	SOYMON037	g1009427	BLASTN	799	1e-57	82
143	4729	700870535H1	SOYMON018	g1009427	BLASTN	405	1e-54	82
144	8117	700752125H1	SOYMON014	g1009428	BLASTN	444	1e-26	75

# SOYBEAN PORPHOBILINOGEN SYNTHASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
145	-700678901	700678901H1	SOYMON007	g493019	BLASTN	1259	1e-101	97
146	-700680455	700680455H1	SOYMON008	g493019	BLASTN	915	1e-105	98
147	-700897467	700897467H1	SOYMON027	g493019	BLASTN	1091	1e-98	97
148	-700994415	700994415H1	SOYMON011	g493019	BLASTN	381	1e-21	97
149	-701002563	701002563H1	SOYMON018	g493019	BLASTN	608	1e-41	96
150	-701208590	701208590H1	SOYMON035	g493019	BLASTN	366	1e-21	92
151	-GM8017	LIB3039-038-Q1-E1-H8	LIB3039	g493019	BLASTN	224	1e-29	86
152	-GM9259	LIB3049-002-Q1-E1-G5	LIB3049	g493019	BLASTN	281	1e-16	84
153	-GM9536	LIB3049-003-Q1-E1-D4	LIB3049	g493019	BLASTN	426	1e-61	82
154	11129	700660017H1	SOYMON004	g313724	BLASTX	176	1e-17	51
155	22115	701208693H1	SOYMON035	g493019	BLASTN	1353	1e-103	97
156	22115	701151960H1	SOYMON031	g493019	BLASTN	1245	1e-94	100
157	22115	700846243H1	SOYMON021	g493019	BLASTN	571	1e-75	99
158	23112	701207084H1	SOYMON035	g493019	BLASTN	1408	1e-108	97
159	23112	700654971H1	SOYMON004	g493019	BLASTN	1236	1e-94	98
160	25460	700656593H1	SOYMON004	g493019	BLASTN	1296	1e-99	99
161	25460	701050212H1	SOYMON032	g493019	BLASTN	1235	1e-94	98
162	25460	701123120H1	SOYMON037	g493019	BLASTN	1245	1e-94	100
163	25460	701055012H1	SOYMON032	g493019	BLASTN	868	1e-93	99
164	3678	LIB3039-036-Q1-E1-D2	LIB3039	g493019	BLASTN	1757	1e-137	99
165	3678	LIB3039-031-Q1-E1-F9	LIB3039	g493019	BLASTN	1736	1e-135	99
166	3678	700553643H1	SOYMON001	g493019	BLASTN	1383	1e-106	98
167	3678	700558620H1	SOYMON001	g493019	BLASTN	1361	1e-104	98
168	3678	701046832H1	SOYMON032	g493019	BLASTN	1185	1e-100	96
169	3678	701109455H1	SOYMON036	g493019	BLASTN	1282	1e-98	97
170	3678	LIB3056-002-Q1-B1-D5	LIB3056	g493019	BLASTN	1286	1e-98	98
171	3678	700844432H1	SOYMON021	g493019	BLASTN	1274	1e-97	99
172	3678	700847337H1	SOYMON021	g493019	BLASTN	1242	1e-94	98
173	3678	700994748H1	SOYMON011	g493019	BLASTN	1221	1e-92	98



174	3678	701213656H1	SOYMON035	g493019	BLASTN	1176	1e-89	99
175	3678	700969539H1	SOYMON005	g493019	BLASTN	1120	1e-88	95
176	3678	700862858H1	SOYMON020	g493019	BLASTN	775	1e-85	98
177	3678	701109689H1	SOYMON036	g493019	BLASTN	968	1e-85	97
178	3678	701105613H1	SOYMON036	g493019	BLASTN	1127	1e-85	98
179	3678	700762772H1	SOYMON015	g493019	BLASTN	1039	1e-84	97
180	3678	LIB3039-047-Q1-E1-D4	LIB3039	g493019	BLASTN	619	1e-82	94
181	3678	700962419H1	SOYMON022	g493019	BLASTN	670	1e-82	99
182	3678	700975590H1	SOYMON009	g493019	BLASTN	965	1e-82	93
183	3678	701108204H1	SOYMON036	g493019	BLASTN	1069	1e-80	99
184	3678	700725416H1	SOYMON009	g493019	BLASTN	633	1e-66	89
185	3678	701055787H1	SOYMON032	g493019	BLASTN	790	1e-57	100
186	3678	700996808H1	SOYMON018	g493019	BLASTN	740	1e-52	100
187	3678	701210127H1	SOYMON035	g493019	BLASTN	545	1e-36	100
188	3678	700739360H1	SOYMON012	g493019	BLASTN	245	1e-21	99
189	3678	700742044H1	SOYMON012	g493019	BLASTN	313	1e-17	98
190	3678	701065288H1	SOYMON034	g493019	BLASTN	335	1e-17	100
191	3678	701110261H1	SOYMON036	g493019	BLASTN	211	1e-12	99

#### SOYBEAN HYDROXYMETHYLBILANE SYNTHASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
154	11129	700660017H1	SOYMON004	g313724	BLASTX	176	1e-17	51
192	-700653648	700653648H1	SOYMON003	g313723	BLASTN	262	1e-60	90
193	-700666293	700666293H1	SOYMON005	g313723	BLASTN	772	1e-65	82
194	-700975534	700975534H1	SOYMON009	g313723	BLASTN	498	1e-76	89
195	-701064190	701064190H1	SOYMON034	g313723	BLASTN	450	1e-62	85
196	-701125566	701125566H1	SOYMON037	g313723	BLASTN	474	1e-55	83
197	11129	700656680H1	SOYMON004	g313723	BLASTN	834	1e-73	85
198	12006	700556506H1	SOYMON001	g313723	BLASTN	652	1e-69	83
199	12006	700557751H1	SOYMON001	g313723	BLASTN	652	1e-65	83
200	12006	700556513H1	SOYMON001	g313723	BLASTN	572	1e-61	83
201	12006	700848102H1	SOYMON021	g313723	BLASTN	459	1e-54	81
202	20966	701108946H1	SOYMON036	g313150	BLASTX	148	1e-13	83
203	20966	701054125H1	SOYMON032	g313150	BLASTX	148	1e-13	83
204	20966	701108239H1	SOYMON036	g313150	BLASTX	122	1e-9	83
205	8428	LIB3052-015-Q1-N1-G5	LIB3052	g313723	BLASTN	865	1e-79	79
206	8428	LIB3055-013-Q1-N1-H6	LIB3055	g313723	BLASTN	1021	1e-76	80
207	8428	701140841H1	SOYMON038	g313723	BLASTN	942	1e-69	85
208	8428	700559220H1	SOYMON001	g313723	BLASTN	871	1e-63	82
209	8428	700998668H1	SOYMON018	g313723	BLASTN	872	1e-63	84
210	8428	701047766H1	SOYMON032	g313723	BLASTN	855	1e-62	85
211	8428	701055336H1	SOYMON032	g313723	BLASTN	848	1e-61	85
212	8428	700558405H1	SOYMON001	g313723	BLASTN	733	1e-57	85
213	8428	700758672H1	SOYMON015	g313723	BLASTN	406	1e-50	85
214	8428	700904365H1	SOYMON022	g313723	BLASTN	392	1e-48	85
215	8428	700987727H1	SOYMON009	g313723	BLASTN	563	1e-38	84
216	8428	701119125H1	SOYMON037	g313723	BLASTN	386	1e-21	84
217	8428	700833610H1	SOYMON019	g313723	BLASTN	314	1e-17	84

SOYBEAN GLUTAMATE-1-SEMIALDEHYDE 2,1-AMINOMUTASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
218	-700659530	700659530H1	SOYMON004	g310566	BLASTN	1138	1e-86	94
219	-700733492	700733492H1	SOYMON010	g310566	BLASTN	552	1e-47	87
220	-700737926	700737926H1	SOYMON012	g310566	BLASTN	421	1e-34	84
221	-700757107	700757107H1	SOYMON015	g310566	BLASTN	640	1e-91	96
222	-700852479	700852479H1	SOYMON023	g310566	BLASTN	1162	1e-88	98
223	-700971066	700971066H1	SOYMON005	g310566	BLASTN	1208	1e-91	96
224	-700982010	700982010H1	SOYMON009	g310566	BLASTN	1016	1e-90	97
225	-700986986	700986986H1	SOYMON009	g310566	BLASTN	1313	1e-100	95
226	-701042351	701042351H1	SOYMON029	g310566	BLASTN	1238	1e-94	99
227	10046	LIB3051-102-Q1-K1-E2	LIB3051	g747967	BLASTN	549	1e-106	90
228	10046	700554665H1	SOYMON001	g310566	BLASTN	981	1e-86	94
229	10046	LIB3040-061-Q1-E11-A4	LIB3040	g310566	BLASTN	441	1e-80	83
230	10046	700560249H1	SOYMON001	g310566	BLASTN	773	1e-70	91
231	10046	700995521H1	SOYMON011	g310566	BLASTN	773	1e-70	91
232	10046	700741513H1	SOYMON012	g310566	BLASTN	608	1e-69	90
233	10046	701207555H1	SOYMON035	g310566	BLASTN	748	1e-68	91
234	10046	701109782H1	SOYMON036	g310566	BLASTN	615	1e-67	90
235	10046	701047870H1	SOYMON032	g310566	BLASTN	493	1e-57	89
236	10046	701108348H1	SOYMON036	g310566	BLASTN	441	1e-51	88
237	10046	701041675H1	SOYMON029	g310566	BLASTN	613	1e-51	89
238	10046	700659465H1	SOYMON004	g310566	BLASTN	407	1e-30	91
239	10046	701144580H1	SOYMON031	g310566	BLASTN	218	1e-9	89
240	11600	700788311H1	SOYMON011	g310566	BLASTN	1234	1e-94	95
241	11600	700902195H1	SOYMON027	g310566	BLASTN	1209	1e-91	96
242	11600	701135233H1	SOYMON038	g310566	BLASTN	1177	1e-89	96
243	12473	701104293H1	SOYMON036	g310566	BLASTN	884	1e-83	93
244	12473	701104392H1	SOYMON036	g310566	BLASTN	754	1e-76	93
245	13619	700877188H1	SOYMON018	g310566	BLASTN	1323	1e-101	99
246	13619	700845619H1	SOYMON021	g310566	BLASTN	948	1e-70	93
247	20047	700660491H1	SOYMON004	g310566	BLASTN	718	1e-67	93
248	20047	700989453H1	SOYMON011	g310566	BLASTN	369	1e-21	95
249	5811	LIB3049-032-Q1-E1-A12	LIB3049	g747967	BLASTN	1188	1e-131	96
250	5811	LIB3049-030-Q1-E1-C2	LIB3049	g747967	BLASTN	1245	1e-98	100
251	5811	701142371H1	SOYMON038	g747967	BLASTN	1209	1e-93	96
252	5811	701155760H1	SOYMON031	g747967	BLASTN	1215	1e-92	100
253	5811	700983730H1	SOYMON009	g310566	BLASTN	417	1e-85	96
254	5811	701064792H1	SOYMON034	g310566	BLASTN	952	1e-81	97
255	5811	700561468H1	SOYMON002	g747967	BLASTN	757	1e-75	90
256	5811	700945376H1	SOYMON024	g310566	BLASTN	884	1e-73	98
257	5811	700756262H1	SOYMON014	g747967	BLASTN	722	1e-72	88
258	5811	700981893H1	SOYMON009	g747967	BLASTN	610	1e-70	92
259	5811	700562668H1	SOYMON002	g747967	BLASTN	807	1e-58	98
260	5811	700979322H1	SOYMON009	g747967	BLASTN	465	1e-54	97
261	5811	700905434H1	SOYMON022	g747967	BLASTN	707	1e-50	97
262	5811	700733377H1	SOYMON010	g747967	BLASTN	640	1e-48	93
263	5811	700562340H1	SOYMON002	g747967	BLASTN	629	1e-43	97
264	5811	701211223H1	SOYMON035	g310566	BLASTN	506	1e-41	94
265	5811	700565140H1	SOYMON002	g310566	BLASTN	416	1e-34	91

## SOYBEAN GLUTAMATE tRNA LIGASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
266	-700653562	700653562H1	SOYMON003	g1008482	BLASTN	280	1e-27	64
267	-700740810	700740810H1	SOYMON012	g2370487	BLASTX	217	1e-22	43
268	-700893754	700893754H1	SOYMON024	g1008483	BLASTX	121	1e-8	34
269	-701009959	701009959H2	SOYMON019	g603849	BLASTX	175	1e-17	97
270	-701011820	701011820H1	SOYMON019	g1322915	BLASTX	129	1e-20	58
271	-701051674	701051674H1	SOYMON032	g2370487	BLASTX	221	1e-23	57
272	-701052937	701052937H1	SOYMON032	g1322915	BLASTX	236	1e-30	69
273	-701060112	701060112H1	SOYMON033	g1322915	BLASTX	121	1e-20	48
274	-701109025	701109025H1	SOYMON036	g157564	BLASTX	179	1e-18	67
275	-GM18124	LIB3065-002-Q1-N1-C2	LIB3065	g2995454	BLASTN	758	1e-66	78
276	-GM36590	LIB3051-050-Q1-K1-D3	LIB3051	g603849	BLASTX	91	1e-37	67
277	20438	700976589H1	SOYMON009	g2370487	BLASTX	253	1e-31	59
278	24353	701054537H1	SOYMON032	g157564	BLASTX	232	1e-24	56
279	24353	701054523H1	SOYMON032	g157564	BLASTX	192	1e-19	70
280	24353	701054530H1	SOYMON032	g157564	BLASTX	195	1e-19	69
281	27156	701137188H1	SOYMON038	g1008483	BLASTX	246	1e-26	53
282	27156	701207661H1	SOYMON035	g1008483	BLASTX	211	1e-24	52
283	27156	700726789H1	SOYMON009	g1008483	BLASTX	182	1e-22	52
284	32173	701202848H1	SOYMON035	g416260	BLASTN	443	1e-54	78
285	32173	LIB3049-020-Q1-E1-G4	LIB3049	g416260	BLASTN	733	1e-52	77
286	32173	700846868H1	SOYMON021	g157564	BLASTX	136	1e-20	67
287	7264	LIB3051-061-Q1-K1-B9	LIB3051	g2995454	BLASTN	841	1e-100	81
288	7712	700666928H1	SOYMON005	g157564	BLASTX	276	1e-30	56
289	7712	700665965H1	SOYMON005	g157564	BLASTX	263	1e-29	55

## SOYBEAN GLUTAMYL-tRNA REDUCTASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
290	-700670004	700670004H1	SOYMON006	g1694925	BLASTN	397	1e-34	75
291	-700728413	700728413H1	SOYMON009	g1049056	BLASTN	696	1e-49	80
292	-700994105	700994105H1	SOYMON011	g1694925	BLASTN	771	1e-55	82
293	-700995896	700995896H1	SOYMON011	g1049057	BLASTX	84	1e-13	86
294	-701099031	701099031H1	SOYMON028	g1015318	BLASTN	459	1e-28	69
295	-701128557	701128557H1	SOYMON037	g1694925	BLASTN	432	1e-27	79
296	-GM35481	LIB3051-036-Q1-K1-G6	LIB3051	g1694925	BLASTN	1070	1e-80	73
297	25545	701123925H1	SOYMON037	g1694925	BLASTN	719	1e-51	78
298	25545	700727539H1	SOYMON009	g1694925	BLASTN	639	1e-44	79
299	2655	700553888H1	SOYMON001	g1694925	BLASTN	242	1e-20	74
300	2655	700553887H1	SOYMON001	g1694926	BLASTX	74	1e-8	93
301	2885	700728635H1	SOYMON009	g1015318	BLASTN	816	1e-59	78
302	2885	701097007H1	SOYMON028	g1015318	BLASTN	706	1e-50	80
303	3203	700986371H1	SOYMON009	g1694925	BLASTN	501	1e-69	80
304	3203	700556832H1	SOYMON001	g1694925	BLASTN	446	1e-47	81
305	3203	700995693H1	SOYMON011	g1039331	BLASTN	390	1e-39	72
306	33811	LIB3051-105-	LIB3051	g1694925	BLASTN	573	1e-36	78

## Q1-K1-B10

## SOYBEAN Mg-CHELATASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
307	-700554488	700554488H1	SOYMON001	g1732468	BLASTN	835	1e-66	85
308	-700657604	700657604H1	SOYMON004	g2318116	BLASTN	975	1e-72	89
309	-700658239	700658239H1	SOYMON004	g1732468	BLASTN	1095	1e-92	100
310	-700737719	700737719H1	SOYMON012	g2318116	BLASTN	973	1e-72	87
311	-700902101	700902101H1	SOYMON027	g1732468	BLASTN	181	1e-15	89
312	-700943788	700943788H1	SOYMON024	g1732468	BLASTN	402	1e-23	91
313	-700992328	700992328H1	SOYMON011	g1732468	BLASTN	856	1e-62	87
314	-700996107	700996107H1	SOYMON018	g1732468	BLASTN	1208	1e-96	95
315	-701050458	701050458H1	SOYMON032	g1732468	BLASTN	881	1e-67	97
316	-701053810	701053810H1	SOYMON032	g1732468	BLASTN	1133	1e-99	99
317	-701119309	701119309H1	SOYMON037	g2318116	BLASTN	944	1e-83	91
318	-701128728	701128728H1	SOYMON037	g1732468	BLASTN	358	1e-28	90
319	-701134728	701134728H2	SOYMON038	g1732468	BLASTN	994	1e-85	95
320	-GM16990	LIB3055-002-Q1-B1-F10	LIB3055	g3059094	BLASTN	1141	1e-86	83
321	-GM18022	LIB3055-011-Q1-N1-A4	LIB3055	g3059094	BLASTN	1079	1e-138	99
322	-GM1974	LIB3028-009-Q1-B1-G12	LIB3028	g3059094	BLASTN	561	1e-66	89
323	-GM22404	LIB3030-009-Q1-B1-A1	LIB3030	g3059094	BLASTN	1104	1e-93	83
324	20034	700944453H1	SOYMON024	g2318116	BLASTN	1089	1e-81	90
325	20034	700834945H1	SOYMON019	g2318116	BLASTN	944	1e-69	87
326	20915	700993045H1	SOYMON011	g1732468	BLASTN	412	1e-23	97
327	20915	700978188H1	SOYMON009	g1732468	BLASTN	400	1e-22	95
328	22044	LIB3028-007-Q1-B1-A8	LIB3028	g2318116	BLASTN	1719	1e-134	89
329	22044	701063842H1	SOYMON034	g2318116	BLASTN	1084	1e-81	92
330	22044	700725383H1	SOYMON009	g2318116	BLASTN	1028	1e-76	90
331	22044	701001558H1	SOYMON018	g2318116	BLASTN	668	1e-75	89
332	26346	LIB3039-010-Q1-E1-G5	LIB3039	g1732468	BLASTN	441	1e-25	84
333	26346	700972265H1	SOYMON005	g1732468	BLASTN	287	1e-12	84
334	26346	701052502H1	SOYMON032	g1732468	BLASTN	241	1e-9	82
335	2822	LIB3054-010-Q1-N1-G10	LIB3054	g1732468	BLASTN	850	1e-120	93
336	2822	LIB3065-011-Q1-N1-D8	LIB3065	g1732468	BLASTN	841	1e-104	83
337	2822	700686645H1	SOYMON008	g1732468	BLASTN	868	1e-82	91
338	2822	700997745H1	SOYMON018	g1732468	BLASTN	878	1e-81	92
339	2822	700994426H1	SOYMON011	g1732468	BLASTN	1084	1e-81	89
340	2822	700739971H1	SOYMON012	g1732468	BLASTN	871	1e-80	95
341	2822	701138070H1	SOYMON038	g1732468	BLASTN	449	1e-65	87
342	2822	701203251H1	SOYMON035	g1732468	BLASTN	605	1e-65	87
343	2822	701206105H1	SOYMON035	g1732468	BLASTN	698	1e-65	88
344	2822	700895378H1	SOYMON027	g1732468	BLASTN	486	1e-62	83
345	2822	701105638H1	SOYMON036	g1732468	BLASTN	739	1e-61	90
346	2822	700898271H1	SOYMON027	g1732468	BLASTN	746	1e-61	89
347	2822	LIB3040-024-	LIB3040	g1732468	BLASTN	652	1e-49	84

Q1-K1-B10 - Q1-K1-B10

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
372	-700839666	700839666H1	SOYMON020	g2623989	BLASTN	736	1e-52	82
373	-700846363	700846363H1	SOYMON021	g439482	BLASTN	732	1e-52	77
374	-700901570	700901570H1	SOYMON027	g2623989	BLASTN	848	1e-61	81
375	-700907558	700907558H1	SOYMON022	g439482	BLASTN	700	1e-49	75
376	-701048026	701048026H1	SOYMON032	g439482	BLASTN	654	1e-45	71
377	-701064702	701064702H1	SOYMON034	g439482	BLASTN	439	1e-26	68
378	-701105159	701105159H1	SOYMON036	g2429617	BLASTN	487	1e-50	77
379	26592	701208376H1	SOYMON035	g439482	BLASTN	722	1e-51	78
380	26592	701097475H1	SOYMON028	g439482	BLASTN	729	1e-51	75
381	26592	701119601H1	SOYMON037	g439482	BLASTN	518	1e-40	78
382	28079	701015447H1	SOYMON019	g439482	BLASTN	840	1e-61	81
383	28079	701102766H1	SOYMON028	g439482	BLASTN	789	1e-56	82

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
384	-700214815	700214815H1	SATMON016	g972938	BLASTX	244	1e-26	80
385	-700222875	700222875H1	SATMON011	g972937	BLASTN	340	1e-27	77
386	-L30662921	LIB3066-008- Q1-K1-C6	LIB3066	g3068702	BLASTN	504	1e-31	69
387	11381	LIB3078-007- Q1-K1-H2	LIB3078	g3068702	BLASTN	281	1e-57	75
388	11381	700084837H1	SATMON011	g972937	BLASTN	281	1e-48	74
389	11381	700088129H1	SATMON011	g972937	BLASTN	317	1e-44	78

390	11381	700045204H1	SATMON004	g972938	BLASTX	363	1e-43	73
391	11381	700084253H1	SATMON011	g972937	BLASTN	317	1e-35	79
392	11381	700427169H1	SATMONN01	g972938	BLASTX	225	1e-33	73
393	11381	700104418H1	SATMON010	g972937	BLASTN	317	1e-26	77
394	17510	700218357H1	SATMON016	g972937	BLASTN	173	1e-12	72
395	17510	700217457H1	SATMON016	g972937	BLASTN	173	1e-11	72
396	1913	700243564H1	SATMON010	g972937	BLASTN	357	1e-38	78
397	1913	700577332H1	SATMON031	g972938	BLASTX	165	1e-15	68

#### MAIZE PROTOCHLOROPHYLLIDE REDUCTASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
398	-700346250	700346250H1	SATMON021	g16117	BLASTN	338	1e-47	86
399	-700430255	700430255H1	SATMONN01	g19060	BLASTN	619	1e-42	93
400	-L30681828	LIB3068-021-Q1-K1-A7	LIB3068	g19060	BLASTN	212	1e-14	69
401	-L30782362	LIB3078-006-Q1-K1-G6	LIB3078	g19060	BLASTN	472	1e-47	73
402	18503	700321674H1	SATMON025	g683475	BLASTN	260	1e-50	82
403	18503	700220580H1	SATMON011	g683475	BLASTN	179	1e-16	80
404	2096	LIB36-005-Q1-E1-G5	LIB36	g683475	BLASTN	1633	1e-127	90
405	2096	LIB3062-018-Q1-K1-E9	LIB3062	g510676	BLASTN	1436	1e-116	84
406	2096	LIB3078-008-Q1-K1-H8	LIB3078	g19060	BLASTN	1017	1e-115	85
407	2096	LIB3078-004-Q1-K1-D7	LIB3078	g19060	BLASTN	1325	1e-101	84
408	2096	LIB3062-010-Q1-K1-G7	LIB3062	g19060	BLASTN	1159	1e-90	81
409	2096	700043426H1	SATMON004	g683475	BLASTN	1189	1e-90	93
410	2096	700093887H1	SATMON008	g19060	BLASTN	1194	1e-90	87
411	2096	700045326H1	SATMON004	g19060	BLASTN	1115	1e-84	92
412	2096	700439164H1	SATMON026	g683475	BLASTN	620	1e-81	89
413	2096	700093546H1	SATMON008	g19060	BLASTN	821	1e-81	88
414	2096	700081986H1	SATMON011	g510676	BLASTN	1076	1e-80	85
415	2096	700044764H1	SATMON004	g683475	BLASTN	1056	1e-79	89
416	2096	700098537H1	SATMON009	g510676	BLASTN	760	1e-73	87
417	2096	700340891H1	SATMON020	g683475	BLASTN	591	1e-70	88
418	2096	700100677H1	SATMON009	g683475	BLASTN	890	1e-65	89
419	2096	700265286H1	SATMON017	g683475	BLASTN	578	1e-53	88
420	2096	700212436H1	SATMON016	g16117	BLASTN	718	1e-51	83
421	2096	700046348H1	SATMON004	g683475	BLASTN	726	1e-51	87
422	2096	700968694H1	SATMONN04	g19060	BLASTN	306	1e-28	87
423	2096	700453783H1	SATMON029	g19060	BLASTN	197	1e-12	78
424	5587	LIB3062-053-Q1-K1-C4	LIB3062	g683475	BLASTN	1366	1e-111	91
425	5587	700087630H1	SATMON011	g683475	BLASTN	1021	1e-91	92
426	5587	700088983H1	SATMON011	g683475	BLASTN	708	1e-88	86
427	5587	700100889H1	SATMON009	g683475	BLASTN	1149	1e-87	89
428	5587	700470729H1	SATMON025	g683475	BLASTN	1006	1e-85	90
429	5587	700100883H1	SATMON009	g683475	BLASTN	839	1e-80	90
430	5587	700214072H1	SATMON016	g683475	BLASTN	910	1e-73	91
431	5587	700044664H1	SATMON004	g683475	BLASTN	587	1e-56	84

432	5587	LIB189-007-Q1-E1-A7	LIB189	g683475	BLASTN	403	1e-53	79
433	5587	700042060H1	SATMON004	g683475	BLASTN	734	1e-52	84
434	5587	LIB83-011-Q1-E1-A8	LIB83	g683475	BLASTN	310	1e-32	74
435	5587	700083223H1	SATMON011	g683475	BLASTN	310	1e-31	79
436	5587	700442503H1	SATMON026	g683475	BLASTN	458	1e-29	89
437	5587	700101540H1	SATMON009	g683475	BLASTN	310	1e-28	78
438	5587	700207961H1	SATMON016	g683475	BLASTN	304	1e-26	82
439	5587	700092960H1	SATMON008	g683475	BLASTN	310	1e-26	72
440	5587	700041761H1	SATMON004	g683475	BLASTN	301	1e-24	86
441	5587	700087935H1	SATMON011	g2244614	BLASTX	141	1e-12	100
442	5632	LIB3069-036-Q1-K1-A5	LIB3069	g510676	BLASTN	849	1e-76	81
443	5632	700243480H1	SATMON010	g16117	BLASTN	806	1e-73	89
444	5632	700198041H1	SATMON016	g16117	BLASTN	738	1e-52	90
445	5632	700097480H1	SATMON009	g19060	BLASTN	712	1e-50	80
446	5632	700088645H1	SATMON011	g19060	BLASTN	659	1e-46	79
447	5632	LIB3068-012-Q1-K1-C6	LIB3068	g19061	BLASTX	129	1e-35	90
448	5632	LIB3069-026-Q1-K1-G11	LIB3069	g19060	BLASTN	496	1e-34	75
449	5632	700081934H1	SATMON011	g16117	BLASTN	519	1e-34	89
450	5632	LIB3062-050-Q1-K1-B3	LIB3062	g16117	BLASTN	348	1e-29	83
451	5632	LIB3069-042-Q1-K1-D6	LIB3069	g19060	BLASTN	404	1e-25	72
452	5632	700095961H1	SATMON008	g19060	BLASTN	277	1e-23	68
453	5632	700091061H1	SATMON011	g19061	BLASTX	86	1e-10	64
454	5632	700224365H1	SATMON011	g19060	BLASTN	213	1e-10	63
455	5632	700089289H1	SATMON011	g19060	BLASTN	213	1e-10	63
456	5632	700094923H1	SATMON008	g19060	BLASTN	183	1e-9	63
457	5632	700094625H1	SATMON008	g19060	BLASTN	206	1e-9	63
458	5632	700093380H1	SATMON008	g19060	BLASTN	206	1e-9	63
459	5632	700093964H1	SATMON008	g19060	BLASTN	213	1e-9	69
460	5632	700083043H1	SATMON011	g19060	BLASTN	193	1e-8	63
461	5632	700095165H1	SATMON008	g510676	BLASTN	195	1e-8	65
462	5633	700094759H1	SATMON008	g510676	BLASTN	629	1e-43	86
463	5633	700094711H1	SATMON008	g16117	BLASTN	468	1e-32	89
464	9949	700213043H1	SATMON016	g16117	BLASTN	1192	1e-90	87
465	9949	700084670H1	SATMON011	g16117	BLASTN	1125	1e-85	88
466	9949	700213929H1	SATMON016	g16117	BLASTN	439	1e-27	84

# MAIZE PUTATIVE PROTOCHLOROPHYLLIDE REDUCTASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
467	-700353742	700353742H1	SATMON024	g348717	BLASTN	512	1e-33	73
468	-700423111	700423111H1	SATMONN01	g348717	BLASTN	684	1e-48	72
469	15163	LIB3069-040-Q1-K1-F9	LIB3069	g348717	BLASTN	874	1e-64	70
470	15163	700623844H1	SATMON034	g348717	BLASTN	752	1e-53	69
471	15163	700623744H1	SATMON034	g348717	BLASTN	725	1e-51	70
472	15163	700623644H1	SATMON034	g348717	BLASTN	661	1e-46	72
473	15163	700612907H1	SATMON033	g348717	BLASTN	597	1e-40	73

474	15163	700612808H1	SATMON033	g348717	BLASTN	579	1e-39	74
475	15163	700623852H1	SATMON034	g348717	BLASTN	489	1e-30	67
476	15163	700475540H1	SATMON025	g348717	BLASTN	413	1e-26	65
477	22562	700571483H1	SATMON030	g348717	BLASTN	447	1e-26	72
478	30690	LIB3062-046-Q1-K1-D4	LIB3062	g348719	BLASTN	514	1e-31	69
479	30690	700425786H2	SATMONN01	g348720	BLASTX	168	1e-17	54

#### MAIZE COPROPORPHYRINOGEN OXIDASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
480	-L30623969	LIB3062-019-Q1-K1-A9	LIB3062	g1213067	BLASTX	123	1e-25	76
481	26808	LIB3062-022-Q1-K1-A9	LIB3062	g1213067	BLASTX	195	1e-35	89
482	26808	LIB36-007-Q1-E1-H7	LIB36	g414665	BLASTN	235	1e-8	85
483	5948	700614009H1	SATMON033	g1212993	BLASTN	1318	1e-101	87
484	5948	LIB3078-027-Q1-K1-C2	LIB3078	g1212993	BLASTN	1185	1e-89	83
485	5948	700207069H1	SATMON003	g1212993	BLASTN	1003	1e-82	81
486	5948	701183985H1	SATMONN06	g1212993	BLASTN	1064	1e-79	88
487	5948	700043235H1	SATMON004	g1212993	BLASTN	944	1e-69	85
488	5948	700237643H1	SATMON010	g1212993	BLASTN	920	1e-67	85
489	5948	700167142H1	SATMON013	g1212993	BLASTN	832	1e-60	85
490	98	LIB3062-011-Q1-K1-B9	LIB3062	g1212993	BLASTN	1515	1e-120	85
491	98	700089965H1	SATMON011	g1212993	BLASTN	1129	1e-85	85
492	98	700473370H1	SATMON025	g1212993	BLASTN	812	1e-79	83
493	98	700018492H1	SATMON001	g1212993	BLASTN	650	1e-45	87
494	98	700336060H1	SATMON019	g1212993	BLASTN	423	1e-26	83

#### MAIZE PROTOPORPHYRINOGEN OXIDASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
495	13987	700397414H1	SATMONN01	g1877018	BLASTX	152	1e-13	72
496	13987	700377840H1	SATMON019	g2370333	BLASTX	115	1e-8	75
497	21128	700087081H1	SATMON011	g1183006	BLASTN	851	1e-62	75
498	21128	700222959H1	SATMON011	g1183006	BLASTN	551	1e-47	74
499	8675	LIB3062-009-Q1-K1-F6	LIB3062	g3093409	BLASTN	1093	1e-82	72

#### MAIZE UROPORPHYRINOGEN DECARBOXYLASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
500	-700210906	700210906H1	SATMON016	g1009429	BLASTX	172	1e-25	50
501	-700334993	700334993H1	SATMON019	g1009427	BLASTN	515	1e-70	84
502	-700432067	700432067H1	SATMONN01	g216564	BLASTX	123	1e-14	39
503	-L1891364	LIB189-002-Q1-E1-E8	LIB189	g1009427	BLASTN	914	1e-78	85
504	-L30625966	LIB3062-056-Q1-K1-D10	LIB3062	g1322019	BLASTX	660	1e-102	100
505	-L30626254	LIB3062-058-Q1-K1-D9	LIB3062	g1009427	BLASTN	516	1e-32	81



506	-L30783694	LIB3078-054-Q1-K1-D9	LIB3078	g1009427	BLASTN	1355	1e-104	84
507	30392	700090031H1	SATMON011	g1009427	BLASTN	794	1e-79	92
508	30392	LIB3062-053-Q1-K1-D9	LIB3062	g1009427	BLASTN	762	1e-54	89
509	30392	LIB3069-027-Q1-K1-G9	LIB3069	g1009427	BLASTN	664	1e-44	83

#### MAIZE PUTATIVE UROPORPHYRINOGEN DECARBOXYLASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
510	-700799143	700799143H1	SATMON036	g48040	BLASTX	128	1e-21	47

#### MAIZE PORPHOBILINOGEN SYNTHASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
511	-700082696	700082696H1	SATMON011	g1041422	BLASTN	544	1e-64	86
512	-700421637	700421637H1	SATMONN01	g1041422	BLASTN	591	1e-57	85
513	10331	700049523H1	SATMON003	g1041422	BLASTN	694	1e-73	84
514	10331	700214149H1	SATMON016	g1041422	BLASTN	986	1e-73	84
515	6252	LIB3060-049-Q1-K1-D11	LIB3060	g1041422	BLASTN	1296	1e-119	85
516	6252	700104193H1	SATMON010	g1041422	BLASTN	1117	1e-84	87
517	6252	700043614H1	SATMON004	g1041422	BLASTN	1081	1e-81	87
518	6252	700104333H1	SATMON010	g1041422	BLASTN	757	1e-76	86
519	6252	700099573H1	SATMON009	g1041422	BLASTN	969	1e-71	85
520	6252	LIB189-034-Q1-E1-G11	LIB189	g1041422	BLASTN	829	1e-64	82
521	6252	700150031H1	SATMON007	g1041422	BLASTN	715	1e-50	80
522	6252	700150305H1	SATMON007	g1041422	BLASTN	494	1e-32	88
523	6664	700098341H1	SATMON009	g1041422	BLASTN	861	1e-62	80
524	6664	700097010H1	SATMON009	g1041422	BLASTN	861	1e-62	80
525	6664	700150830H1	SATMON007	g1041422	BLASTN	655	1e-45	78
526	6664	700088427H1	SATMON011	g1041422	BLASTN	598	1e-41	85
527	6664	700216648H1	SATMON016	g1041422	BLASTN	586	1e-40	78
528	6664	700150750H1	SATMON007	g1041422	BLASTN	562	1e-38	77
529	6664	700089504H1	SATMON011	g1041422	BLASTN	349	1e-35	81
530	6664	700150781H1	SATMON007	g1041422	BLASTN	473	1e-30	71
531	6664	700071849H1	SATMON007	g1041423	BLASTX	158	1e-14	66

#### MAIZE HYDROXYMETHYLBILANE SYNTHASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
532	-700042853	700042853H1	SATMON004	g2661765	BLASTN	1319	1e-101	96
533	-700209530	700209530H1	SATMON016	g2661765	BLASTN	1046	1e-96	91
534	-L30784536	LIB3078-039-Q1-K1-D10	LIB3078	g2661765	BLASTN	980	1e-73	81
535	18	700434552H1	SATMONN01	g2661765	BLASTN	819	1e-59	77
536	18	700621233H1	SATMON034	g2661765	BLASTN	606	1e-44	85
537	18	700621333H1	SATMON034	g2661765	BLASTN	607	1e-41	88
538	22370	LIB3078-049-Q1-K1-D11	LIB3078	g2661765	BLASTN	1197	1e-91	93
539	22370	LIB3078-007-Q1-K1-F2	LIB3078	g313723	BLASTN	745	1e-72	71

540	22370	700223478H1	SATMON011	g313723	BLASTN	712	1e-50	72
541	22370	700216196H1	SATMON016	g313723	BLASTN	508	1e-45	74
542	22370	700551081H1	SATMON022	g2661765	BLASTN	328	1e-36	91

#### MAIZE GLUTAMATE-1-SEMIALDEHYDE 2,1-AMINOMUTASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
543	-L841669	LIB84-026-Q1-E1-D6	LIB84	g506383	BLASTX	164	1e-42	69
544	11095	LIB3078-054-Q1-K1-E2	LIB3078	g556018	BLASTN	1229	1e-93	81
545	11095	LIB83-005-Q1-E1-B11	LIB83	g556018	BLASTN	1169	1e-88	80
546	11095	700101450H1	SATMON009	g556018	BLASTN	1024	1e-76	81
547	11095	700342512H1	SATMON021	g556018	BLASTN	995	1e-74	82
548	11095	700265085H1	SATMON017	g556018	BLASTN	996	1e-74	82
549	11095	700154602H1	SATMON007	g556018	BLASTN	491	1e-58	82
550	11095	700154908H1	SATMON007	g556018	BLASTN	722	1e-58	81
551	11095	700017624H1	SATMON001	g556018	BLASTN	775	1e-55	83
552	11095	700018108H1	SATMON001	g556018	BLASTN	758	1e-54	84
553	11095	700443671H1	SATMON027	g556018	BLASTN	662	1e-46	73
554	11095	700442812H1	SATMON026	g556018	BLASTN	578	1e-39	80
555	11095	700343762H1	SATMON021	g556018	BLASTN	565	1e-38	80
556	11095	700094251H1	SATMON008	g19873	BLASTX	167	1e-16	89
557	11225	LIB3060-054-Q1-K1-C12	LIB3060	g556018	BLASTN	817	1e-69	77
558	11225	700100123H1	SATMON009	g556018	BLASTN	787	1e-56	85
559	11225	700405062H1	SATMON027	g556018	BLASTN	453	1e-34	75
560	11225	700219159H1	SATMON011	g556018	BLASTN	328	1e-26	74
561	11225	700209352H1	SATMON016	g506383	BLASTX	174	1e-17	70
562	11225	700053276H1	SATMON008	g506383	BLASTX	131	1e-10	96
563	11225	700156122H2	SATMON007	g506383	BLASTX	120	1e-9	100
564	15553	700084357H1	SATMON011	g556018	BLASTN	1104	1e-83	80
565	15553	700441108H1	SATMON026	g556018	BLASTN	1071	1e-80	86
566	15553	700441006H1	SATMON026	g556018	BLASTN	1062	1e-79	86
567	15553	700087059H1	SATMON011	g556018	BLASTN	421	1e-26	82
568	20096	700089246H1	SATMON011	g556018	BLASTN	474	1e-49	82
569	20096	700171369H1	SATMON013	g556018	BLASTN	560	1e-37	78

#### MAIZE GLUTAMATE tRNA LIGASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
570	-700614160	700614160H1	SATMON033	g31958	BLASTX	113	1e-10	56
571	-L1892744	LIB189-012-Q1-E1-F8	LIB189	g31958	BLASTX	146	1e-28	50
572	-L1894036	LIB189-029-Q1-E1-B1	LIB189	g157564	BLASTX	143	1e-28	62
573	12385	LIB3067-058-Q1-K1-H9	LIB3067	g2995455	BLASTX	416	1e-70	69
574	13776	700344387H1	SATMON021	g157564	BLASTX	108	1e-19	57
575	21786	700221143H1	SATMON011	g157564	BLASTX	287	1e-32	59
576	26250	LIB3069-031-Q1-K1-E6	LIB3069	g2995455	BLASTX	166	1e-43	74
577	3350	LIB3069-025-	LIB3069	g157564	BLASTX	232	1e-44	46

		Q1-K1-F6						
578	3350	700072785H1	SATMON007	g157564	BLASTX	249	1e-26	45
579	3350	700049536H1	SATMON003	g157564	BLASTX	227	1e-24	50
580	3350	700077013H1	SATMON007	g157564	BLASTX	232	1e-24	49
581	3350	700209830H1	SATMON016	g157564	BLASTX	210	1e-22	52
582	3350	700168681H1	SATMON013	g157564	BLASTX	156	1e-14	40
583	5345	LIB3059-036-Q1-K1-G10	LIB3059	g2995455	BLASTX	148	1e-28	67
584	9230	LIB143-053-Q1-E1-G8	LIB143	g31958	BLASTX	341	1e-55	58
585	9230	700331892H1	SATMON019	g157564	BLASTX	162	1e-31	55

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
586	-700094403	700094403H1	SATMON008	g1039331	BLASTN	740	1e-63	83
587	-700151003	700151003H1	SATMON007	g1039331	BLASTN	885	1e-64	87
588	-700167046	700167046H1	SATMON013	g1039331	BLASTN	772	1e-64	89
589	-L30661635	LIB3066-003-Q1-K1-A8	LIB3066	g1666078	BLASTN	298	1e-18	77
590	-L30661878	LIB3066-012-Q1-K1-F3	LIB3066	g2967440	BLASTN	485	1e-29	85
591	-L362024	LIB36-016-Q2-E2-H11	LIB36	g2920319	BLASTN	170	1e-9	70
592	22014	700045741H1	SATMON004	g1039331	BLASTN	921	1e-67	82
593	22014	700214783H1	SATMON016	g1039331	BLASTN	860	1e-62	83
594	22618	700086081H1	SATMON011	g1039331	BLASTN	1064	1e-79	82
595	22618	700104481H1	SATMON010	g1039331	BLASTN	955	1e-70	81
596	22618	700356789H1	SATMON024	g1039331	BLASTN	644	1e-44	86
597	30084	LIB3062-026-Q1-K1-H5	LIB3062	g2920319	BLASTN	1043	1e-78	87
598	30084	701179026H1	SATMONN05	g1039331	BLASTN	753	1e-72	88
599	6787	LIB36-021-Q1-E1-D9	LIB36	g1039331	BLASTN	1281	1e-97	87
600	6787	701163632H1	SATMONN04	g1039331	BLASTN	941	1e-79	86
601	6787	700162337H1	SATMON012	g1039331	BLASTN	901	1e-66	84
602	6787	700100879H1	SATMON009	g1039331	BLASTN	608	1e-65	86
603	6787	700425112H1	SATMONN01	g1039331	BLASTN	230	1e-9	81
604	9690	LIB3078-023-Q1-K1-F12	LIB3078	g1039331	BLASTN	1755	1e-137	88
605	9690	700097404H1	SATMON009	g1039331	BLASTN	1311	1e-100	89
606	9690	700099954H1	SATMON009	g1039331	BLASTN	1132	1e-92	90
607	9690	700213724H1	SATMON016	g1039331	BLASTN	1211	1e-92	89
608	9690	700468009H1	SATMON025	g1039331	BLASTN	1187	1e-90	88
609	9690	700042554H1	SATMON004	g1039331	BLASTN	931	1e-68	88

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
610	-700042626	700042626H1	SATMON004	g861198	BLASTN	489	1e-31	67
611	-700045010	700045010H1	SATMON004	g861198	BLASTN	824	1e-59	80
612	-700046489	700046489H1	SATMON004	g861198	BLASTN	246	1e-11	93
613	-700090155	700090155H1	SATMON011	g2239151	BLASTX	226	1e-24	84
614	-700100867	700100867H1	SATMON009	g2239150	BLASTN	944	1e-69	81

615	-700152555	700152555H1	SATMON007	g861198	BLASTN	800	1e-57	83
616	-700166615	700166615H1	SATMON013	g861198	BLASTN	438	1e-27	88
617	-700214027	700214027H1	SATMON016	g847872	BLASTN	683	1e-67	80
618	-700216016	700216016H1	SATMON016	g861198	BLASTN	408	1e-24	78
619	-700219682	700219682H1	SATMON011	g2318116	BLASTN	616	1e-42	70
620	-L30606220	LIB3060-019-Q1-K1-D2	LIB3060	g861199	BLASTX	106	1e-32	67
621	15513	700379674H1	SATMON021	g2239151	BLASTX	138	1e-11	87
622	15984	700223402H1	SATMON011	g2318116	BLASTN	830	1e-60	78
623	15984	701185568H1	SATMONN06	g2318116	BLASTN	787	1e-56	79
624	15984	701185572H1	SATMONN06	g2239150	BLASTN	528	1e-48	78
625	15984	700257978H1	SATMON017	g2239150	BLASTN	454	1e-33	72
626	19005	700025578H1	SATMON004	g861198	BLASTN	874	1e-63	84
627	19005	700442062H1	SATMON026	g861198	BLASTN	226	1e-23	81
628	19969	700100921H1	SATMON009	g847872	BLASTN	1183	1e-89	87
629	19969	700422986H1	SATMONN01	g847872	BLASTN	598	1e-75	88
630	19969	700216568H1	SATMON016	g847872	BLASTN	822	1e-59	89
631	21239	LIB36-010-Q1-E1-H9	LIB36	g847872	BLASTN	1338	1e-139	88
632	21239	700053384H1	SATMON009	g847872	BLASTN	596	1e-85	88
633	21239	700043650H1	SATMON004	g847872	BLASTN	1122	1e-84	90
634	21239	700160759H1	SATMON012	g847872	BLASTN	760	1e-74	89
635	29840	LIB84-024-Q1-E1-B8	LIB84	g861198	BLASTN	1195	1e-90	80
636	29840	700046445H1	SATMON004	g861198	BLASTN	898	1e-65	82
637	29840	700099676H1	SATMON009	g861198	BLASTN	755	1e-54	79
638	29840	700216064H1	SATMON016	g861198	BLASTN	433	1e-25	74
639	3221	700342738H1	SATMON021	g2239150	BLASTN	678	1e-74	82
640	3221	700090050H1	SATMON011	g2239150	BLASTN	556	1e-66	80
641	3221	700259555H1	SATMON017	g2239151	BLASTX	215	1e-27	72
642	5373	LIB3078-054-Q1-K1-A2	LIB3078	g861198	BLASTN	1220	1e-92	82
643	5373	700097131H1	SATMON009	g861198	BLASTN	1071	1e-80	82
644	5373	700043508H1	SATMON004	g861198	BLASTN	1017	1e-75	83
645	5373	700046404H1	SATMON004	g861198	BLASTN	988	1e-73	83
646	5373	700431104H1	SATMONN01	g861198	BLASTN	769	1e-65	82
647	5373	700043839H1	SATMON004	g861198	BLASTN	747	1e-53	83
648	5373	700433984H2	SATMONN01	g861198	BLASTN	493	1e-32	76
649	5953	LIB3078-052-Q1-K1-A9	LIB3078	g861198	BLASTN	1148	1e-92	81
650	5953	700045450H1	SATMON004	g861198	BLASTN	1005	1e-74	83
651	5953	700041735H1	SATMON004	g861198	BLASTN	969	1e-71	83
652	5953	LIB83-001-Q1-E1-F5	LIB83	g861198	BLASTN	706	1e-58	83

## MAIZE FERROCHELATASE

Seq No.	Cluster ID	CloneID	Library	NCBI gi	Method	Score	P-value	%Ident
653	-700214704	700214704H1	SATMON016	g2429617	BLASTN	822	1e-59	85
654	-700239147	700239147H1	SATMON010	g439480	BLASTN	905	1e-66	87
655	-700382669	700382669H1	SATMON024	g439480	BLASTN	1079	1e-81	87
656	-700441938	700441938H1	SATMON026	g2429617	BLASTN	766	1e-55	84
657	-700579576	700579576H1	SATMON031	g439480	BLASTN	739	1e-52	75
658	-L1892866	LIB189-014-	LIB189	g2429617	BLASTN	248	1e-15	78

659	-L832454	Q1-E1-E6 LIB83-005-Q1- E1-F4	LIB83	g2460250	BLASTN	303	1e-35	88
660	11690	700151225H1	SATMON007	g2429617	BLASTN	1083	1e-81	91
661	11690	700106040H1	SATMON010	g2429617	BLASTN	627	1e-59	90
662	11690	700167395H1	SATMON013	g2429617	BLASTN	637	1e-44	91
663	14766	LIB143-007- Q1-E1-D5	LIB143	g439480	BLASTN	580	1e-56	77
664	14766	700263637H1	SATMON017	g439480	BLASTN	259	1e-10	72
665	14766	LIB36-007-Q1- E1-C9	LIB36	g2460250	BLASTN	243	1e-8	72
666	16136	700354263H1	SATMON024	g439480	BLASTN	987	1e-76	87
667	16136	700222207H1	SATMON011	g439480	BLASTN	903	1e-66	82
668	17054	700158001H1	SATMON012	g439480	BLASTN	1091	1e-82	90
669	17054	700157069H1	SATMON012	g439480	BLASTN	1096	1e-82	90
670	17054	700453159H1	SATMON028	g439480	BLASTN	816	1e-73	88
671	394	700622934H1	SATMON034	g439480	BLASTN	763	1e-84	86
672	394	700621391H1	SATMON034	g439480	BLASTN	775	1e-82	86
673	394	700098357H1	SATMON009	g2460251	BLASTX	155	1e-13	83
674	9731	LIB3078-007- Q1-K1-B10	LIB3078	g2429617	BLASTN	1299	1e-123	85
675	9731	700355823H1	SATMON024	g2429617	BLASTN	826	1e-84	87
676	9731	700208201H1	SATMON016	g2429617	BLASTN	702	1e-64	82
677	9731	700167772H1	SATMON013	g2429617	BLASTN	648	1e-45	88



#### **\*Table Headings**

##### **Cluster ID**

A cluster ID is arbitrarily assigned to all of those clones which belong to the same cluster at a given stringency and a particular clone will belong to only one cluster at a given stringency. If a cluster contains only a single clone (a “singleton”), then the cluster ID number will be negative, with an absolute value equal to the clone ID number of its single member. The cluster ID entries in the table refer to the cluster with which the particular clone in each row is associated.

##### **Clone ID**

The clone ID number refers to the particular clone in the PhytoSeq database. Each clone ID entry in the table refers to the clone whose sequence is used for (1) the sequence comparison whose scores are presented and/or (2) assignment to the particular cluster which is presented. Note that a clone may be included in this table even if its sequence comparison scores fail to meet the minimum standards for similarity. In such a case, the clone is included due solely to its association with a particular cluster for which sequences of one or more other member clones possess the required level of similarity.

##### **Library**

The library ID refers to the particular cDNA library from which a given clone is obtained. Each cDNA library is associated with the particular tissue(s), line(s) and developmental stage(s) from which it is isolated.

##### **NCBI gi**

Each sequence in the GenBank public database is arbitrarily assigned a unique NCBI gi (National Center for Biotechnology Information GenBank Identifier) number. In this table, the

0933318-012059

NCBI gi number which is associated (in the same row) with a given clone refers to the particular GenBank sequence which is used in the sequence comparison. This entry is omitted when a clone is included solely due to its association with a particular cluster.

### **Method**

The entry in the “Method” column of the table refers to the type of BLAST search that is used for the sequence comparison. “CLUSTER” is entered when the sequence comparison scores for a given clone fail to meet the minimum values required for significant similarity. In such cases, the clone is listed in the table solely as a result of its association with a given cluster for which sequences of one or more other member clones possess the required level of similarity.

### **Score**

Each entry in the “Score” column of the table refers to the BLAST score that is generated by sequence comparison of the designated clone with the designated GenBank sequence using the designated BLAST method. This entry is omitted when a clone is included solely due to its association with a particular cluster. If the program used to determine the hit is HMMSW then the score refers to HMMSW score.

### **P-Value**

The entries in the P-Value column refer to the probability that such matches occur by chance.

### **%Ident**

The entries in the “%Ident” column of the table refer to the percentage of identically matched nucleotides (or residues) that exist along the length of that portion of the sequences which is aligned by the BLAST comparison to generate the statistical scores presented. This entry is omitted when a clone is included solely due to its association with a particular cluster.